



Decision Support for the Rolling Stock Dispatcher

Groth, Julie Jespersen

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Groth, J. J. (2009). *Decision Support for the Rolling Stock Dispatcher*. Technical University of Denmark. DTU Compute PHD

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Decision Support for the Rolling Stock Dispatcher

Julie Jespersen Groth

Kongens Lyngby 2008
IMM-PHD-2008-212

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

IMM-PHD: ISSN 0909-3192

Summary

Real-time recovery is receiving a fast growing interest in an increasingly competitive railway operation market. This thesis considers the area of rolling stock dispatching which is one of the typical real-time railway dispatching problems. All work of the thesis is based on the network and planning processes of the railway operator DSB S-tog a/s.

In the thesis the problems existing in the railway planning process from the strategic to real-time level are briefly sketched. Network planning, line planning, timetabling, crew and rolling stock planning is outlined and relevant references are given. Specifically the thesis references the operation research studies based on the railway operation of DSB S-tog a/s. Subsequently the process of dispatching is outlined with a specific emphasis on rolling stock.

The rolling stock recovery problem is the problem of assigning train units to train departures in a disrupted rolling stock schedule so that operation returns quickly to the originally planned schedule. Different network structures and mathematical formulations for the problem are discussed. Based on prior work on network structures a decomposed approach for the rolling stock recovery problem is put forward.

The main contributions of the thesis are contained in four papers included as appendices. The papers deal with respectively an analysis of robustness in timetables, the mathematical model behind a decision support tool for reinsertion of a train line, a survey on the dispatching problems of passenger railway transportation and the decomposed solution process of the rolling stock recovery problem.

The paper on the robustness analysis has been accepted for submission in the International Journal of Operations Research. Two of the papers have been submitted to journals and are being reviewed. The last paper will be submitted. Furthermore, the work of the two papers on the robustness analysis respectively the reinsertion model have formed the basis of practical projects in DSB S-tog. The applicability of the decomposed process will be further investigated in the future.

Resumé

Genopretning i real tid tiltrækker i dag mere og mere interesse i et jernbanemarked med øget konkurrence. Denne afhandling omhandler området materieldisponering, som er et af de typiske problemer i realtids jernbanedisponering. Alt arbejde i afhandlingen er baseret på DSB S-tog a/s netværk og planlægningsprocesser.

I afhandlingen skitseres kort problemstillingerne i planlægningsprocessen fra strategisk til realtids niveau i jernbanedrift. Netværksplanlægning, linieplanlægning, køreplanslægning, materiel- og mandskabsplanlægning opridses og relevante referencer gives. Specifikt forsøger afhandlingen at referere operations analyse forskning baseret på jernbanedriften som den er hos DSB S-tog a/s. Efterfølgende bliver processen omkring disponering og specielt materieldisponering opridset.

Ved uregelmæssigheder i driften forårsaget af foreksempel nedbrud af tog eller signalfejl er det ikke muligt at fortsætte driften ud fra den oprindelige materielplan. Problemet vedrørende materielgenopretning består i at tildele togsæt til togafgange i en revideret plan på en sådan måde, at driften i en sådan situation hurtigt returnerer til den oprindelige materielplan. Forskellige netværksopstillinger og matematiske formuleringer diskuteres. Baseret på arbejdet med netværksopstillinger bliver foreslået en løsningsmetode for materielgenopretningsproblemet baseret på dekomposition.

Hovedbidragene i denne afhandling er indeholdt i fire artikler inkluderet som appendiks. Rapporterne omhandler henholdsvis en analyse af robusthed i køreplaner, den matematiske model bag et beslutningsstøttværktøj til genindsættelse af en toglinie, et overblik over disponeringsproblemstillingerne for passager-jernbanetransport,

og løsningsmetoden for materielgenopretningsproblemet.

Rapporten om robusthedsanalysen er accepteret til publicering i tidsskriftet International Journal of Operations Research. To af rapporterne er indsendt til tidsskrifter og under bedømmelse. Den sidste rapport vil også blive indsendt. Derudover har arbejdet bag de to projekter om henholdsvis robusthedsanalyse og genindsættelsesmodel dannet grundlaget for praktiske projekter i DSB S-tog. Anvendelsesmulighederne for den dekomponerede løsningsmetode vil blive undersøgt yderligere i fremtiden.

Preface

This thesis was prepared at DTU Informatics and DTU Management, the Technical University of Denmark in partial fulfillment of the requirements for acquiring the Ph.D. degree in engineering.

The thesis focuses on the area of rolling stock dispatching and related decision support methods. The methods and results are based on the real-life plans and operation of DSB S-tog a/s, a railway operator in the greater suburban area of Copenhagen. DSB S-tog a/s is part of the DSB Group. The Ph.D. project has been supervised by associated professor Jesper Larsen. The industrial supervisor is professor Jens Clausen who is part time employed as a chief analysts at DSB S-tog A/S.

The thesis consists of a summary report and a collection of four research papers written in the period 2005–2008.

Lyngby, October 2008

Julie Jespersen Groth

Papers included in the thesis

- [B] Julie Jespersen Groth, Daniel Potthoff, Jens Clausen, Dennis Huisman, Leo Kroon, Gábor Maróti, and Morten Nyhave Nielsen. Disruption Management in Passenger Railway Transportation. IMM-Technical Report-2007-03
- [C] Mads Hofman, Line Frølund Madsen, Julie Jespersen Groth, Jens Clausen and Jesper Larsen. Robustness and Recovery in Train Scheduling - a Simulation Study from DSB S-tog a/s. Accepted for publication in Computers & OR, special issue for Nordic MPS 2006. IMM-Technical Report-2006-12
- [D] Julie Jespersen Groth, Jens Clausen and Jesper Larsen. Optimal reinser-tion of cancelled train line. IMM-Technical Report-2006-13
- [E] Julie Jespersen Groth, Jesper Larsen and Jens Clausen. The rolling stock recovery problem. IMM-Technical Report-2008-18

Acknowledgements

First of all I would like to thank my supervisor and my industrial supervisor for ongoing support and inspiration.

A special thanks to my colleague analysts in production planning, DSB S-tog for productive sparring.

Also, I would like to thank the dispatchers at DSB S-tog for their sincere interest in the tools I can create for them. And for letting me sit in on their watches so that I could gain understanding of what the operation really is seen from the rolling stock perspective.

I would like to thank Professor R. Ahuja for the many discussion we had on network structures and problem formulations.

Most importantly I would like to thank my husband for his love and support. And my brother for being there for our family when I couldn't.

x

Contents

Summary	i
Resumé	iii
Preface	v
Papers included in the thesis	vii
Acknowledgements	ix
1 Introduction	1
1.1 Rolling stock dispatching and recovery	3
1.2 Main contributions of thesis	4
1.3 Outline of thesis	5
2 DSB S-tog	7
2.1 The organization DSB S-tog a/s	7

2.2	Characteristics of DSB S-tog	8
3	Railway planning	15
3.1	Rolling stock planning	18
3.2	Conclusion	25
4	Railway dispatching and recovery	27
4.1	Disruptions	28
4.2	Recovery in practice	29
4.3	Conclusion	33
5	Rolling stock recovery - discussing network structures	35
5.1	Problem description	35
5.2	Discussion of network structures	38
5.3	A decomposed approach	43
5.4	Conclusion	48
6	Discussion of thesis papers	49
6.1	Disruption management in passenger railway transportation . . .	50
6.2	Robustness and recovery in train scheduling - a simulation study from DSB S-tog a/s	51
6.3	Optimal reinsertion of cancelled train line	53
6.4	Rolling stock recovery problem	54
7	A heuristic approach for the RSRP	57

7.1	A construction heuristic	57
7.2	Improvement heuristic	62
7.3	Conclusion	64

8

Future research projects

65

8.1	Comlumn generation approach to the RSRP	65
8.2	Rolling stock depot recovery	67
8.3	Evaluation of robustness of a RSR solution by simulation	68

9

Conclusion

71

9.1	Main contributions	71
9.2	Further developments	73
9.3	Summing up	73

A

Generalized Label Setting Algorithm

75

B

Disruption Management in Passenger Railway Transportation

77

B.1	Introduction	79
B.2	Description of disruption management	82
B.3	Timetable adjustments	92
B.4	Rolling stock re-scheduling	99
B.5	Crew re-scheduling	104
B.6	Integrated Recovery	109
B.7	Conclusions	110

C	Robustness and Recovery in Train Scheduling - a simulation study from DSB S-tog a/s	117
C.1	Background	118
C.2	Related work	121
C.3	Recovery strategies	122
C.4	Background of the problem	125
C.5	Assumptions	128
C.6	The simulation model	130
C.7	Test Cases	133
C.8	Computational Results	135
C.9	Extensions and concrete application	142
C.10	Conclusions and future work	144
D	Optimal reinsertion of cancelled train line	147
D.1	Introduction to DSB S-tog	148
D.2	The Reinsertion Problem	152
D.3	The Reinsertion Model	157
D.4	Computational Experience	160
D.5	An Improved Planning Process	162
D.6	Further Developments	163
D.7	Conclusion	163
E	The rolling stock recovery problem	167
E.1	Introduction	168

E.2 Literature review	170
E.3 Basic elements of a rolling stock plan	172
E.4 Defining a disruption	175
E.5 Network description	178
E.6 The train unit position model	179
E.7 Train sequence model	183
E.8 The train Routing model	185
E.9 Computational results	189
E.10 Conclusion	208

CHAPTER 1

Introduction

In the European railway industry there has been a decline in market shares since the seventies, Pietrantonio and Pelkmans [2004]. To even out or to turn the decline, the European commission increased focus on the general area of European railway transportation in the middle of the nineties. In 1996 a white paper was composed, which is titled "Strategy for Revitalizing the Community's Railways", EU Commision 1996. The intention was to encourage railway operators to cut their costs, improve the quality of service and offer new products. All later initiatives to legislations concerning the further development of railway industry in Europe originate from this white paper¹. The increased focus on improving the operations of railway industry lead to an increased interest in doing railway related research. One of the theoretical areas experiencing increased activity in railway related research is Operations Research. Focus has been on several areas including timetabling, crew and rolling stock planning and train routing. Many of the planning problems have been formulated using integer programming formulations and solved by different optimization techniques, including both exact and heuristic approaches.

In this thesis we will concentrate on rolling stock planning, specifically from the real-time perspective. Our research is based on the practical problems of the suburban railway operator DSB S-tog a/s (S-tog).

¹See the EU commission home page for further information, <http://ec.europa.eu/transport/rail/overview>

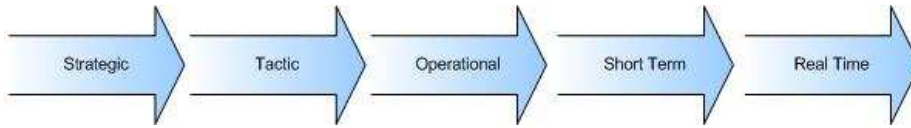


Figure 1.1: The different planning phases seen with respect to time

The research within rolling stock planning of self-propelled rolling stock started back in 1993 where Schrijver [1993] considered the tactical rolling stock planning problem. Since then most focus has been on the tactical and operational planning phases. The last years the results within these phases have matured to be applicable in practice.

Many aspects in rolling stock planning are repeated at the different planning levels. Making a tactic rolling stock plan share the structure of operational rolling stock planning meaning that the same steps are considered at both planning phases. Figure 1.1 shows the different planning phases seen from a time perspective. The longest long-time planning perspective is called *strategic planning*. It spans questions with a time horizon on more than a year e.g. ordering of new rolling stock units. A year prior to operation the planning perspective is called the *tactic planning* level. In this phase the yearly, public timetable is constructed. At S-tog also preliminary rolling stock and crew plans are decided upon. At the *operational planning* level, 2-3 month before operation, the timetable, the rolling stock and crew plans are adjusted according to planned rail maintenance works. When the *short term planning* level is reached from approx. 14 days to 1 day before operation the final adjustments are made to the rolling stock and crew plans. Also, the specific drivers and train units are assigned to departures in the timetable. Finally at the *real-time planning* level, adjustments are made according to departures not yet covered by crew and rolling stock and action is taken in the case of disruption. The latter is referred to as recovery.

My personal motivation for working with the problem area of rolling stock dispatching is a direct consequence of my employment as an operations researcher for the Copenhagen suburban railway operator DSB S-tog (S-tog). I have at S-tog specifically worked with the area of rolling stock planning at strategic, tactic and real-time levels. Significant to the areas of real time rolling stock planning is that so far, no decision support exists for the different real-time rolling stock problems.

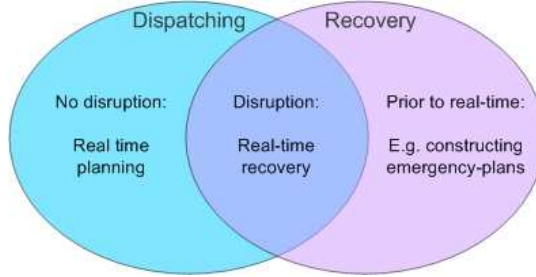


Figure 1.2: Illustrating the interplay of the areas of dispatching and recovery

1.1 Rolling stock dispatching and recovery

The main area of this thesis is rolling stock dispatching and especially the area of Rolling Stock Recovery (RSR). Rolling stock dispatching refers to all planning of rolling stock conducted during operation i.e. in real time.

The areas of dispatching and recovery are two interrelated and interdependent areas concerning real-time planning, see Figure 1.2. We describe the area of dispatching as planning in real-time i.e. all planning carried out during day-of-operation. Dispatching can be pre-planned i.e. each morning a certain planning task must be conducted. Dispatching can also be the planning suddenly being necessary due to disruptions in the operation. The dispatching conducted here belongs to the area of recovery. Recovery concerns all planning of "what to do when disruptions occur". Recovery is conducted in real time as a part of the dispatching task. Before day-of-operation recovery concerns e.g. preplanning actions in case certain parts of the network is blocked. For example, if one side of the tracks in a double tracked network is suddenly blocked for traffic, the trains must be re-routed so that no collisions occur. Plans for timetable, rolling stock and crew can be made in advance for how to handle such a situation.

We divide the real-time planning in railway operation into four main areas, *Train routing*, *Rolling stock recovery*, *Rolling stock depot recovery* and *crew recovery*. The process of recovery is often complicated further due to the fact that different groups of dispatching personnel have the responsibility of recovery within each of these main recovery areas.

The process of train routing is the process of dispatching the departures of the timetable. E.g. given a specific railway network with a set of departures: If a departure is delayed so much that it is no longer possible to conduct the planned operation, choices must be made on which departures to operate, cancel, move

or swap. The objective is quickly to return to the originally planned, public timetable.

The tasks of rolling stock recovery can be divided into two phases. First train unit types are assigned to each terminal departure according to the estimated passenger demand of the departure. The order of train unit types assigned to each departure must also be considered. When the number and order of train unit types are known for each departure in the recovery time-window the physical train units must be routed i.e. a line of works is made for each train unit. The duration of the recovery process from recovery start until the return to the original plan is an important objective of rolling stock recovery. Given a solution to the rolling stock recovery problem it must preferably be able to meet the planned end-of-day balance of rolling stock on each rolling stock depot. If the end-of-day balance is met, dispatchers will be able to start up operation as planned the next morning i.e. the disruption is not spread over several days.

Related to and dependant of rolling stock recovery is recovery of the rolling stock depot plan. For each rolling stock depot a schedule exists for the parking of train units during the day i.e. train units arriving at the depot are parked at specific depot tracks and are leaving the depot on specific times to cover specific departures. A disrupted rolling stock plan will possibly also disrupt the rolling stock depot plan. Rolling stock depot recovery is the process of recovering the rolling stock depot plans according to a recovered rolling stock plan.

The task of reinserting a cancelled train line in the operation concerns the choice of which departures to reinsert, naturally the task would lie within the area of train routing. However, at S-tog the rolling stock dispatcher is responsibility of finding a reinsertion scheme of a train line.

Disruption in general and recovery of the timetable will often affect the crew plan. The recovery process of assigning train drivers to train departures in real time is referred to as crew recovery.

1.2 Main contributions of thesis

This thesis represents one of the first contributions regarding real-time decision support for rolling stock dispatchers. This is supported by Huisman et al. [2005] in which a recent survey is supplied on all the planning phases from strategic to real time. The first submitted contribution is supplied by Nielsen [2008].

Enclosed in the thesis are four papers.

The first paper is a survey paper of the area of disruption management in passenger railway transportation. The aim of the paper is mainly to give a clear description of what disruption management is in the context of passenger railway transportation and to give an overview of the work conducted in the area.

The second paper is on the area of robustness. A simulation model has been developed to test the robustness of different timetables i.e. the timetables ability to absorb disruptions is tested. Also, three different recovery methods are tested for their ability to recover after disruption.

In the third paper an integer programming model is described for optimal reinsertion of a cancelled train line. Cancelling a train line is one of the recovery methods tested in the second paper. When the operation is once again running without delay the train line is reinserted. The process of reinsertion is determined by the rolling stock dispatcher. The model for optimal reinsertion of a cancelled train line is the base of a decision support tool developed for the rolling stock dispatcher and in use in S-tog.

The fourth and final paper describes a decomposed approach for general rolling stock recovery. The rolling stock recovery problem is that of allocating rolling stock to train departures in such a way that each train departure is covered sufficiently according to expected passenger demand and so that the train departures assigned to each individual rolling stock unit form a legal work path. The goal for the decomposed model is that it may form a basis of a decision support system for the rolling stock dispatcher.

1.3 Outline of thesis

After this introductory chapter the railway company S-tog is introduced in Chapter 2. The organizational structure and relevant characteristics of S-tog are described.

In Chapter 3 the area of railway planning is described. We will go through the areas of network planning, line planning, timetabling, crew planning, and rolling stock planning. Especially the area of rolling stock planning will receive our attention.

The specific area of railway dispatching is addressed in Chapter 4. The interaction between the various areas of dispatching is described. Again, the area of rolling stock dispatching is treated in greater detail describing larger and smaller

theoretical problems within the area.

We address the rolling stock recovery problem in Chapter 5. Specifically we discuss different problem representations and network structures.

The papers included in the thesis are discussed further in Chapter 6. A work in progress is presented in Chapter 7. The work in progress concerns a heuristic approach for the rolling stock recovery problem.

Chapter 8 presents and discusses a set of future projects related to rolling stock dispatching.

Chapter 9 contains the conclusion and an overview of the main contributions of the project.

CHAPTER 2

DSB S-tog

The work of this thesis has its primary outset in the railway network and planning of the Danish railway operator S-tog.

2.1 The organization DSB S-tog a/s

S-tog is owned by the Railway group DSB which mainly operates in Denmark. S-tog is an independent railway company of DSB operating the suburban railway network of the Copenhagen outer and inner city area.

BaneDanmark (BD) is the infrastructure manager owning and administrating the physical railway network in Denmark. BD is owned by the Danish State.

In 2007 the yearly turnover of S-tog was approx. 2.8 bill. dkk. The average daily level of passengers is around 240,000 counting both weekdays and weekends.

The activities of S-tog is subject to a contract¹ with the Danish Ministry of Transport stating the number of departures that must be offered to the passengers in the suburban area of Copenhagen. This contract also states the service

¹The complete contract is available online, see of Transport [2004].



Figure 2.1: Organization of DSB S-tog a/s

level that must be offered to the passenger. Service level is measured in the level of punctuality and reliability that should be reached in operation, and in the average number of seats relative to the expected demand.

S-tog has a standard organizational structure, see 2.1. The responsibility of the operation lies with the Production Division. Here, all planning of rolling stock and crew is carried out in the Production Planning department.

2.2 Characteristics of DSB S-tog

2.2.1 Networks and depots

Typical for the S-tog network is that the structure of the timetable is periodic with a frequency of twenty minutes on each line. The network has a tree structure which limits the number of possible connections in the network, see Fig. 2.2. In the planning process this feature often reduces the size of the planning problems or makes them easier to formulate and solve.

Many of the ideas of planning are adaptable to larger railway networks, however, the planning problems handled in larger railway networks are larger and may have another structure e.g. not completely periodic or with network-cycles.

The network of S-tog has 84 stations, all sections in the network are covered

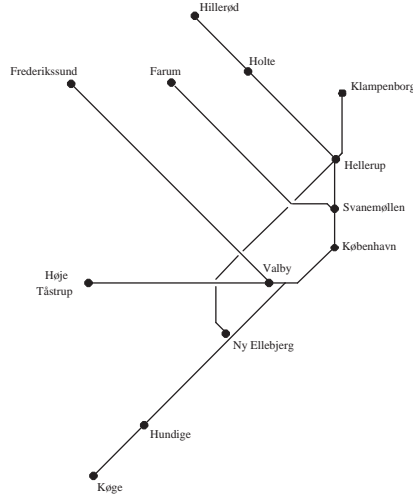


Figure 2.2: Network of DSB S-tog a/s

by double tracks except a small section of 500 meters near one of the section terminals. The double tracks are normally directed meaning that trains keep to the right of their direction of travel. In the case of rail maintenance work the direction of tracks may be changed.

Rolling stock depots are located at section terminals and at some other station used as train line terminals. All S-tog depots are relatively small, see Tab. 2.1. All maintenance of rolling stock takes place in the workshop located in Høje Taastrup. Expanded cleaning of rolling stock, such as removal of graffiti, is performed in Hundige where a train wash is located. The cleaning facilities in Hundige we will refer to as the prepare center.

2.2.2 Trains and train lines

A *train line* is defined by two terminals and a predefined stopping pattern for trains on the line between the terminals. Each train line has a number of *trains* that together form a closed circuit with a constant frequency. For example, given a train line with a route that can be driven back and forth between the terminals in 110 minutes, including turnaround time, and a 20 minutes frequency the train line will have 6 trains. The train line structure of the public timetable of S-tog is illustrated in Fig. 2.3. In Fig. 2.4 the line plan for the total network is

Depot	Track length (approx. in meters)
Køge	~900
Hundige	~2500
København	~3000
Klampenborg	~800
Hillerød	~1500
Farum	~500
Frederikssund	~1000
Ballerup	~900
Høje Taastrup	~4000
(Holte)	~300

Table 2.1: Available track at rolling stock depots

illustrated as it was in 2007. A terminal departure is defined by start/end times and start/end terminals. A *train sequence* for each train indicate the *terminal departures* of the train. The terminal departures are also called train tasks. Hence, a train sequence consists of a sequence of train tasks. Rolling stock or *train units* are assigned to the terminal departures during the day.

The rolling stock schedule can be viewed from two different perspectives. A *train unit perspective* and a *train line perspective*. Both are types of Gantt charts. In the train unit perspective each line in the Gantt chart is equivalent to a train unit i.e. in the line, the train departures of the train units are illustrated, see Fig. 2.5. In the train sequence perspective there is a line in the Gantt chart for each train in the train line. The train units assigned to the train tasks of the train sequence is illustrated as blocks in the Gantt chart, see Fig. 2.6. Fig. 2.5 and 2.6 illustrate the exact same fractions of a rolling stock plan. At S-tog the rolling stock schedule is always viewed from the *train line* perspective.

S-tog has at present two different rolling stock types or *train unit types*, see Fig. 2.7 and 2.8. We will refer to them as SE, the shorter train unit, and SA, the longer train unit type. Tab. 2.2 shows some data on the two train unit types. Train units can be combined to form larger or smaller *train compositions*. A train composition is some ordered combination of train units consisting of one to three train units. Legal train compositions are listed in Tab. 2.3. In practice mostly train compositions consisting of up to two units are used. In the planning process the preferred composition length is no more than two train units.

We define the north and south end of a composition in the following way. No where in the S-tog network can train units turn around to face in the opposite

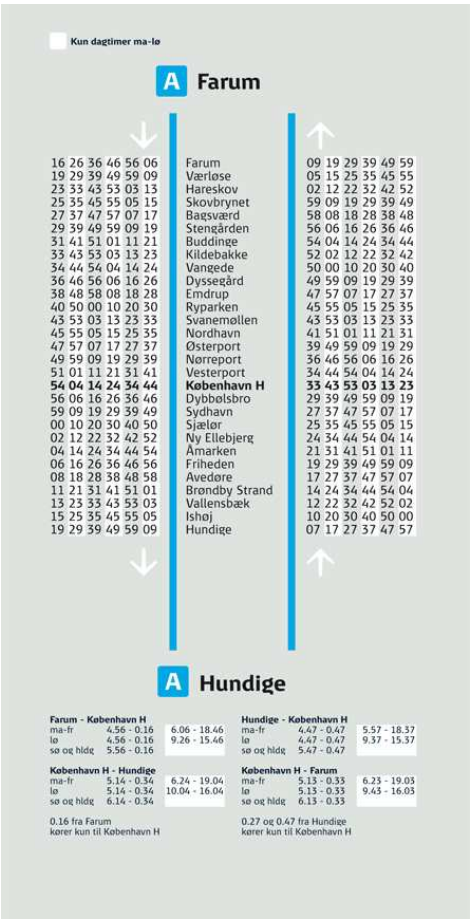


Figure 2.3: The public timetable anno 2008 for line A

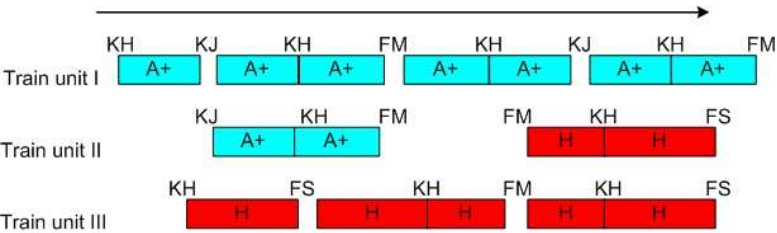


Figure 2.5: The rolling stock schedule seen from the train unit perspective. KH , KJ and FM are terminal stations. $A+$ and H are specific train lines.

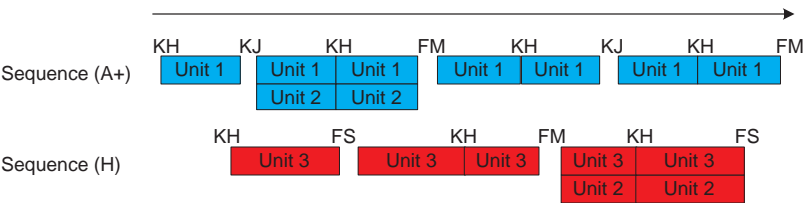


Figure 2.6: The rolling stock schedule seen from the train sequence perspective. KH , KJ and FM are terminal stations. $A+$ and H are specific train lines.

direction. The northern end of a train unit is geographically in the north when the train unit is at the Copenhagen central station.

Data	SE	SA
Length in meters	46	86
Cars	4	8
Seat capacity	150	336

Table 2.2: Data on train unit types

Operation can be quantified by train kilometers and train unit kilometers. Train kilometers are the distance the train drives within a certain time. Train unit kilometers are train kilometers multiplied by the number of train units put on a train.

Composition
SE
SA
SE - SE
SE - SA
SA - SE
SA - SA
SE - SE - SE
SE - SE - SA
SE - SA - SE
SA - SE - SE

Table 2.3: All legal train compositions



Figure 2.7: SE, the short train unit types



Figure 2.8: SA, the long train unit types

CHAPTER 3

Railway planning

This chapter introduces railway planning from the railway operator perspective and some relevant references for each planning problem. For a comprehensive survey of passenger railway optimization models, see Caprara et al. [2007]. We will first discuss the problems of the planning process, see Fig. 3.1. Secondly, we consider the area of rolling stock planning in more detail.

Network planning refers to the area of planning expansions of the railway network. Network planning is in Denmark conducted by the National Rail Authorities in cooperation with the infrastructure manager. The time perspective is the strategic level with up to 20 years time horizon. Each rail operator maintains a list, in the form of a argumentative report, of suggestions for expanding the network. This prioritized list forms part of the rail operators contribution to the political discussion of where to expand or close down tracks. There is to our knowledge no research within this area of planning, probably because most of the decisions here are based on political and economical considerations.

Based on the existing network, line planning is conducted. Recall that the public timetable of S-tog has a train line structure, see Chapter 2. In the line planning process it is decided how many train lines are needed to cover the network, which stations will be terminals and approximate stopping patterns for each train line, and finally what the frequency will be. Borndörfer et al. [2008] supply the most recent survey on line planning in public transportation

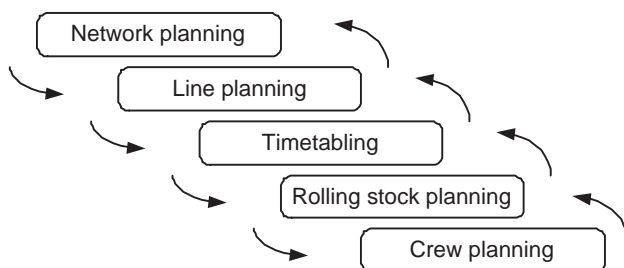


Figure 3.1: Illustrating the different railway planning problems

in general. Two Ph.D. theses, Bussieck [1998] and Goosens [2004], contain descriptions of various models for railway line planning. Recent work includes Goosens et al. [2004] which considers a branch-and-cut approach for two formulations of the Cost-Minimizing Line-Planning Problem and Bussieck et al. [2004] which presents a heuristic based on both a linear and a nonlinear formulation. Bussieck et al. [2004] address line planning in general, however, tests are based on a rail example. Most recently, Borndörfer et al. [2007] suggest a new multi-commodity flow model for line planning, which takes a dynamic approach to developing lines. They present a solution method and show test results for the German city of Potsdam for line planning of trams and busses. So far no work has been started to develop system support to the line planning process in S-tog. The S-tog line plan is generally based on considerations of overall demand on the different network sections.

Most railway operations run according to timetables. At S-tog, the timetable is changed annually in agreement with the infrastructure manager. As mentioned in Chapter 2 the timetable is periodic. A lot of research has been done within the area of periodic timetabling. The first contribution was by Serafini and Ukovich [1989]. Lately two survey chapters, Liebchen [2007] and Liebchen and Möhring [2007] were published. The most recent contribution is Kroon et al. [2008]. The paper describes a stochastic model for distributing buffer times in timetables. A Master thesis project and a Bachelor project has been made which focus particularly on the timetabling at S-tog, see Villumsen [2006] and Andersen et al. [2006]. Also, Nielsen et al. [2006] present a mixed integer programming model for timetabling developed and used by S-tog.

In practice there is an increasing interest in using automatic support when evaluating the potential of constructed timetables. Often referenced systems for evaluating timetables are e.g. PETER which is based on Max-Plus algebra, see Goverde and Odijk [2002], or SIMONE, a simulation tool used in the Netherlands, see Middelkoop and Bouwman [2001]. In Hofman et al. [2006] timetables are evaluated using a macroscopic level simulation model. The timetables are based on the S-tog network and differ widely in basic structure. Also, a simulation model based on the railway simulation system RailSys exists for the S-tog network. The model has been used for timetable evaluations as well as evaluation of various rail network extensions.

On the basis of the timetable, schedules for rolling stock and crew are created. At S-tog this is also an annually process. The schedules are modified during the year according to changes caused by e.g. track maintenance work.

We partition crew in train drivers and other groups of personnel. In particular we will describe different planning phases concerning the train drivers and then mention some of the S-tog related research concerning planning of ticket inspectors.

The strategic planning concerning crew is mainly related to when to hire new drivers and how many to hire. In Folkmann et al. [2007] two mixed integer programming model are presented. The latter model can be used for estimating the future need for train drivers given a low level detailed timetable and a set of low level detailed duty templates. The templates give approximate schedules of when the train drivers are available for driving, having breaks etc.

Also, there may be some strategic crew planning involved in locating or relocating crew depots. The crew depot location problem involves assessment of the expenses of the location of each depot with respect to number of drivers needed to run a day of operation and the expected punctuality and reliability of the operation given the location of depots. To our knowledge, no work has been published on the crew depot location problem.

Tactic crew planning concerns constructing the specific duties. This process is called either duty scheduling or crew pairing. A duty is a line of work or a set of train tasks put together in such a way that they form a feasible work plan for a driver for one day. Feasibility refers in this context to the work rules agreed upon with the unions. After the duties have been constructed they are grouped in rosters. A roster is a set of duties grouped in such a way that a driver can cover the roster feasibly. The process of constructing rosters is called rostering.

At the tactic level S-tog uses the planning system TURNI supplied by the Italian Software company **Double-Click** for the crew pairing problem. See Kroon and

Fischetti [2001] for a case from the Dutch railway company NS Reizigers. In Nielsen and Christensen [2006] the authors describe how a statistical Design of Experiments were used to fine tune TURNI for use in practice at S-tog, see Montgomery [1997] for a comprehensive introduction to the area of Design of Experiments. Extensive published research material is available on tactic crew planning. A recent railway-related survey is supplied by Bengtsson et al. [2007] on the crew pairing problem. A survey on rostering is supplied by Kohl and Karisch [2004]. At S-tog the rostering process is done manually in groups representing the drivers.

The operational crew planning is much like the tactical crew planning. The operational crew planning level in S-tog consist of deciding and implementing changes to the tactic crew plan. No system support is available for this process. However, presently projects run at S-tog evaluating the use of TURNI as an applicable tool for the process.

At the real time level only little published work is available, see Huisman et al. [2005]. An older contribution on the crew recovery problem is available in Walker et al. [2004]. In this paper the authors present a joint model for crew recovery and real time train rerouting. A recent paper on crew recovery is presented in Rezanova and Ryan [2006]. The paper presents a set partitioning formulation solved with a LP relaxation and Branch-and-Price approach. The research is based on real case data from S-tog. S-tog has subsequently started a project with the purpose of implementing the work of Rezanova and Ryan [2006] in practice.

Research within crew recovery in the airline industry is more well-established than that of rail. For a general survey on airline disruption management, see Kohl et al. [2007].

Related to crew planning is the planning of the ticket inspectors, see Nielsen [2006]. The objective of this problem is to cover departures by ticket inspectors. Not all departures can be covered. It is also an objective to keep some variability in the overall schedule for the inspectors.

3.1 Rolling stock planning

In this section we discuss rolling stock planning at all planning levels from strategic level to real time.

3.1.1 Strategic rolling stock planning

The main task at the strategic planning level is to decide how many train units are necessary in the future. In Folkmann et al. [2007] a mixed integer programming model is presented that based on an approximate demand-curve for the morning peak, which is the time of day most likely to be impacted by delays, covers trains on the circuit running in the time period. Different objectives are used to get a clear impression on how many new train units must be obtained of each train type in question in order to still being able to cover the operation. The results has in S-tog been used together with economic analysis to estimate future rolling stock needs.

3.1.2 Tactic and operational rolling stock planning

Tactic and operational rolling stock planning are very similar in problem structure.

As mentioned S-tog yearly constructs a rolling stock schedule according to the timetable. This schedule is the tactic rolling stock plan. It defines the composition to be assigned to each timetable departure. Recall, a composition is a set of train units of specific train types in a specific order. Given as input is the timetable and an expected passenger demand for each timetable departure. A tactic rolling stock schedule exists for each day type of the week i.e. one for each weekday Monday to Friday, one for Saturday and one for Sunday. The schedule is kept cyclic from day to day and from week to week. That is, given the balance of train units on rolling stock depots at the end of each Monday, a standard operation for Tuesdays can be started on Tuesday morning. Also, the balances on Sundays enables the typical Monday start the next morning.

The operational problem of S-tog consist of adjusting the tactical rolling stock plan according to changes at shorter term caused e.g. by rail maintenance work.

The tactic and operational problem of assigning rolling stock to train tasks is thoroughly explored, see Huisman et al. [2005] for a joint survey. Here, it is called the rolling stock circulation problem. The problem is to find a proper allocation of rolling stock units to the trips (train tasks). There is a distinction between the problems of allocating rolling stock when the fleet is composed by train units and when it is composed by train carriages and train locomotives. Papers concerning the former problem are Schrijver [1993], Ben-Khedher et al. [1998], Peeters and Kroon [2003], Alfieri et al. [2006] and Fioole et al. [2006]. Papers concerning the latter problem are Brucker et al. [2004], Cordeau et al.

[2001] and Lingaya et al. [2002]. We will concentrate mainly on the problem of determining rolling stock circulations for self-propelling multiple train units. The adjective multiple refers to the train unit being able to operate in multiple with other units of same type.

A system is currently being developed by DSB and DSB S-tog based on the theoretical models and results of Fioole et al. [2006], which is based upon the work described in Alfieri et al. [2006] and Peeters and Kroon [2003].

Below, the papers relevant to our problem are described, that is, these papers concentrate on the tactical respectively operational planning of rolling stock. Hence, the objective is to find cost efficient, virtual "work plans" for rolling stock.

Schrijver [1993]: Minimum circulation of railway stock

In this paper the author determines a minimum circulation of rolling stock on a single train line running from Amsterdam to Vlissingen and vice versa. For each train the expected passenger demand indicate how many train units of either first or second class are necessary. The objective is to ensure a sufficient number of seats available in each train task. The model does not take the train unit order within a composition into account.

In the case of only one train unit type the author presents a minimum-cost circulation model for which the optimal solution is guaranteed integral. The minimum-cost circulation formulation is solved both by applying a standard algorithm for this type of model based on min-cost augmenting paths and cycles and by applying a linear programming solver, CPLEX. Running times are in both cases 0.05 CPU seconds.

In the case of two train units types the model becomes a multi-commodity circulation model. The running time when solving this with CPLEX is several hours. Therefore, tightening techniques are deployed so that constraints for the train types are tightened to describe exactly the convex hull of the solutions in the solution space polygon. For the considered instance there are 99 polygons corresponding to each train task in the problem, they can be tightened within 0.04 CPU seconds in a preprocessing phase of the solution approach. Also, the author applies an order by which the branch-and-bound procedure invoked by CPLEX is selecting variables. After implementation of these two techniques, CPLEX provides a solution within 1.58 CPU seconds.

Ben-Khedher et al. [1998]: Schedule Optimization at SNCF: From Conception to Day of Departure

In this paper the authors discuss the problem of capacity adjustment based on the problem of finding railway capacity for high speed trains running in the TGV network of SNCF, France. The model is based on the seat reservation system and the objective is to maximize expected profit.

Alfieri et al. [2006], Efficient Circulation of Railway Rolling Stock

The problem of constructing circulations of train units is addressed in this paper. Focus is on a single line. The model couples and decouples train units from trains as the depots are passed. The objective is to minimize the cost of allocating train units to departures. The order within each composition is taken into consideration. The model is tested for two train types.

The problem is a complex multi-commodity flow problem where the order of commodities matter. The network of the problem is a time line network with an aggregation of consecutive arrivals and departures at the same station. The inequalities of the model are tightened by adding valid inequalities.

The composition order is handled by using a transition graph for each train. In this type of graph the nodes are possible compositions for a certain train task. Arcs represent feasible transitions between compositions on consecutive tasks.

The solution approach is based on a hierarchical decomposition into sub problems. First, the model, not taking compositions into consideration, is solved. Second, it is checked whether there is a feasible solution for the composition problem. The latter is done by solving a series of subproblems, where each subproblem corresponds to fixing the variables with respect to each task and composition type.

The results include tests on a single line with different objective functions.

Peeters and Kroon [2003]: Circulation of Railway Rolling Stock: A Branch-and-Price Approach

A branch-and-price algorithm for solving the allocation of train units to a single line or a set of interacting train lines is presented. Input is a periodic timetable and the passenger demand. The model is tested on several real-life instances of the railway operator, NS Reizigers.

Objectives considered are those of minimizing the seat shortage, minimizing the number of shunting operations and the number of driven train unit kilometers. Constraints considered are flow conservation constraints and, inventory constraints.

The model is based on a transition graph as is the model described in Alfieri et al. [2006].

The authors apply a Dantzig-Wolfe decomposition, reformulating the problem so that a variable is associated with each path through the transition graph of all trains. The problem is divided into a Restricted Master Problem (RMP) and a column generation model generating columns for the RMP. Integer solutions are obtained using branch-and-price.

Fioole et al. [2006]: A rolling stock circulation model for combining and splitting of passenger trains

This paper presents a model for finding the compositions of train units on train tasks. Each solution is feasible with respect to composition order in depots and with respect to depot capacities. The model additionally takes into consideration combining and splitting of trains in depot junctions. It is an extension of the model described in Peeters and Kroon [2003].

The model is a multi-commodity flow model. The objective considers minimizing with respect to efficiency, service and robustness. The objective term of efficiency means minimization of driven train unit kilometers, service means minimizing the number of standing passengers and robustness means minimizing the number of (de-/)coupling operations on depots.

After presenting the model, the authors present an aggregation of the model containing more variables but fewer constraints.

The model is implemented and solved in CPLEX. The authors apply various speed-up techniques; adjusting parameters of CPLEX, using priorities when branching, and exploiting the structure of the instances at hand. The problems is run for 2 hours. Results indicate the optimality gap achieved within the 2 hour limit and the time it takes to reach the best result. The latter vary between 1900 and 7100 CPU seconds.

A heuristic is also applied. First a feasible solution is found searching the neighborhood of the LP optimum. Afterwards, a local search heuristic is applied. This procedure improves the solution up to 6 % with respect to number of driven train unit kilometers.

Abbink et al. [2005]: Allocation of Railway Rolling Stock for Passenger Trains

The problem presented in this paper is slightly different from the problems presented in the papers described so far. The model does not consider circulations

of rolling stock. It concentrates solely on allocating train units to trains running in the 8 o'clock hour. The 8 o'clock hour is the heaviest loaded morning rush hour with respect to number of passengers which is why this period is chosen.

The model minimizes the train shortage km during rush hours to optimality. It is implemented in OPL Studio 3.1 and solved with Cplex 7.0. It has been tested on several scenarios based on 2001-2002 timetables of NS Reizigers.

3.1.3 Short term rolling stock planning

At the short term planning level there are still a few adjustment to the plans constructed at tactic and operational level e.g. if the weather is warmer than normal trains going to Klampenborg (where one of the popular Copenhagen beaches are) on Saturdays must be as large as possible. Due to weather forecasts this is first known a few days in advance.

Also, at the short term level physical train units are assigned to departures. It is assured in this process that maintenance constraints are satisfied so that no train unit enters a circuit where it will not be able to get back to the maintenance workshop within required kilometer limits. This routing of rolling stock so that the train units are in runs that passes the workshop in time with respect to the rolling stock maintenance requirements is also called the *Maintenance routing problem*. Two recent papers present two different models for the problem, see Maróti and Kroon [2007] and Maróti and Kroon [2005].

Related to the planning of the operation is maintenance planning at the workshop. The maintenance planning at the workshop relates to scheduling the regular maintenance checks and sudden occurring defects at the workshop according to how the trains arrive at the workshop. The problem is basically a job shop scheduling problem, see Jain and Meeran [1999] for a survey. Two master theses have contributed to the research within workshop planning for S-tog, see Jensen [2008] and Jacobsen [2008].

3.1.4 Planning at rolling stock depots

Related to the rolling stock planning is the planning of the rolling stock depots. At the rolling stock depots train units are stored when they are not running in operation, at the workshop or at the prepare center. The rolling stock depots consist of a set of parallel tracks of different lengths where the train units are parked. At S-tog the depot tracks are only open in one end. This implies that

train units can only access and leave a depot track in a last in - first out fashion.

Train units enter and leave the depot during the day according to how train units are decoupled and coupled to trains. There is no time available during the operation to rearrange train units. Therefore, the train units must be parked at the depot tracks in such a way that the right train unit is always available in the open end of a depot track when it has to leave the depot.

A further challenge for the planning of S-tog depots is that they are all with two exceptions small in size. At the rolling stock depot in Farum there is e.g. only track space for up to 6 SA train units, see 2.1.

In the master thesis by Føns [2006] a solution approach meant for decision support is presented. It is based on a S-tog case. The research is inspired by the representation found in the article by Freling et al. [2005]. For comprehensive descriptions and discussions of the relative few works within the depot planning area, see Føns [2006].

3.1.5 Real time rolling stock dispatching

In this project we address train unit allocation in real time. There are strong parallels between rolling stock allocation solved on the tactical level (the normal plan) and the rolling stock recovery problem. The overall objectives are the same; to minimize the cost of covering the timetable with rolling stock and to cover the demand as given by estimated number of passengers on a departure. The main difference between the tactical and the real time levels is that in real time, work plans for actual train units are made, and at the tactical level train runs (lines of work for anonymous train units) are constructed without specific train units being allocated to these.

The real time problem input specify how many kilometers each train unit has driven. For each train task assigned to a train unit is then checked that the kilometers already driven plus the kilometers corresponding to the task do not exceed the kilometer limit for next maintenance check. In the same way, an upper limit on the number of days between maintenance checks exists. Finally, defects categorized into different levels of severity further limit the time before the train unit must return to the workshop.

The overall problem is to construct lines of work for train units for a given time period. A line of work consists of train tasks given by a departure time and place, and an arrival time and place. The lines of work must fulfil a number of constraints. It is the objective to cover all train tasks with as few train unit

kilometers as possible and at the same time supply the best possible coverage of the passenger demands given for each train task.

A related reference to the rolling stock recovery problem is Nielsen [2008]. In this paper a generic framework for the rolling stock re-scheduling problem is presented. The problem described concerns the re-balancing of rolling stock on train tasks in real time. The problem is considered at train type level. The framework presented is an expansion of the model in Fioole et al. [2006] where constraints are included considering the end-of-day balances of rolling stock. The objectives taken into consideration are the number of cancelled trips, changes to the original rolling stock plan and changes to the end-of-day off balances on rolling stock depots. The model is solved with CPLEX 10.1. The registered computation times varies from few seconds up to a minute depending on the problem instances solved. All computational results are based on data from the Dutch railway operator NS Reizigers.

3.2 Conclusion

In this chapter we have presented the planning phases of a rail operator. We have focused on the European perspective and based our descriptions on the planning practices of the suburban railway operator S-tog. For a recent review on Operations Research within passenger railway operation in the United States, see Ahuja et al. [2005a].

We have concentrated on the railway operator related planning which is concentrated on the planning of the timetable, the resources of rolling stock and crew and related planning of e.g. shunting at the rolling stock depots. For a recent, comprehensive introduction and review of infrastructure manager related planning, which is planning relating to the utilization of the rail network see D'Ariano [2008].

The main planning problems illustrated in Fig. 3.1 are solved and the result or output of the planning problem is past on as input to the next planning problem. With no automatic planning support only one main solution is found for each problem, however, minor adjustments may be carried out to an preceding problem if the plan next in the field becomes too expensive, non-robust or infeasible. For example, some departures may be adjusted in the timetable, if this implies better turn-over times for crew at the crew depot. When automatic planning systems exist for a planning problem several solutions are evaluated, but again only minor adjustments will be made to solutions of problems earlier in the planning process.

CHAPTER 4

Railway dispatching and recovery

This chapter discusses the topic of railway dispatching and recovery. As described in chapter 1 dispatching and recovery are two closely related areas. Whereas dispatching is referring to all the actions taking in real time, recovery refers to the dispatching made in the case of a disrupted situation to get back to the originally planned schedule. Recovery can refer to decisions made in real time or they can refer to recovery strategies developed prior to the operation.

The infrastructure manager has the responsibility for the safe processing of departures and thereby also the right to dispatch the timetable departures during the operation. The rail operator has the responsibility of dispatching the resources available for the departures, i.e. rolling stock and crew. The dispatching of resources is directly dependent on the dispatching of timetable departures i.e. the change in train routing directly affects the states of rolling stock and crew. A detailed description on passenger railway transportation disruption management and all involved parties is given in Groth et al. [2007].

4.1 Disruptions

Delays occur in the operation. The delays occurring may have varying effect on the timetable and the underlying resource schedules of crew and rolling stock.

4.1.1 Incidents

We partition incidents which cause delay in two main categories, primary and secondary incidents.

Primary incidents are a consequence of external factors. In this case external means something which is not the timetable structure itself. True external incidents could be delays caused by passengers, accidents in the railway network or heavy snow. Others (not true) external incidents occurring in practice are delays caused by failure in infrastructure, sudden rolling stock defects or train drivers being late for work.

Primary incidents on rolling stock can be caused by defects on train units e.g. a defect windscreen wiper in rainy weather, which implies that the train unit must not drive. Primary incidents can also be caused by the recovery decisions of the network traffic controller e.g. the network traffic controller can choose to turn a train on line A early to a train on line E. This means that the A-train take over the train tasks of the E-train and that the train units of the composition of the A-train now are on the E-train. Possibly these train units will then end up at a wrong end depot in the evening, if they stay assigned to the E-train. Hence, a disruption has occurred.

As for the rolling stock the primary incidents related to the crew plan are incidents concerning drivers not being able to cover their assigned train task or timetable recovery decisions made by the rolling stock dispatcher directly affecting the crew plan. An example of the first is a driver signing in too late thereby not being able to cover his first train task. An example of the second could be the same situation as mentioned for rolling stock. When the A-train is turned to perform the train tasks of the E-train the train driver will then be on the "wrong" train and very likely he will not be able to cover the next train task assigned to him in his planned duty. Hence, a disruption has occurred.

Secondary incidents are a consequence of one or several incidents which are primary or secondary. Secondary delays are also called knock-on-delays. These are a direct effect of the way the timetable is constructed. For example, let us consider two consecutive departures in a rail network, δ_1 and δ_2 where δ_1

departs τ minutes before δ_2 . If δ_1 is delayed by μ minutes and $\mu > \tau$, δ_2 is also delayed. That is, the delay of δ_1 also affects δ_2 .

Knock-on-delays on rolling stock are caused by the delays of the timetable delaying the arrival and departure times of train tasks. This in it self does not create a disruption in the timetable unless the delay of a present train task t is so large that a train unit can not cover its next train task, $\nu(t)$. In this situation another train unit may be reinserted to cover $\nu(t)$

Similar to the secondary delays on rolling stock the knock-on-delays occurring for crew are caused by the delays of the timetable. A train task may be delayed to an extent so that the train driver assigned to the train task, t , will be late for covering the next train task in his duty.

4.1.2 Effect of delays on the planned operation

We categorize delays according to their severity. Minor delays are delays, which can be absorbed through buffers in the timetable. Large delays have a severe impact on the timetable and the resource schedules. A large delay is the result of e.g. a temporary blockage of the network. In between minor and large delays are all those delays that cannot be absorbed by the timetable but with some recovery action can be reduced so that the operation can return to normal. Often this type of "middle-size" delays is a consequence of several minor delays occurring within a short period of time, see Hofman et al. [2006].

4.2 Recovery in practice

When an incident occurs, which disrupts the timetable and perhaps the crew or rolling stock schedules, a recovery process is started. Each of the dispatching groups take action within their own field. The network traffic controller makes the overall decisions which particularly concerns recovery of the timetable departures. The initial disruption or the recovery choices made by the network traffic controller will most likely affect the resource schedules of crew and/or rolling stock.

In the S-tog operation there are today practically no decision support systems used which can automatically generate and suggest recovery solutions. Instead recovery situations are supported by incorporating robustness in schedules and preplanned recovery strategies constructed before the day of operation. These

are used together with ad hoc solutions decided by each individual dispatcher.

Incorporated robustness in schedules: Robustness is incorporated into the schedules by distributing buffer times in them. Buffer times can be included in the stopping time at stations, turning times at terminals and the running time between pairs of stations. Buffer times are included in all plans to reduce the amount of knock-on-delays and thereby increase the robustness of the plan. A study on how to distribute buffer times in a timetable is available in Kroon et al. [2008].

Preplanned recovery rules: There exist schedules made prior to the operation with the purpose of reducing the number of departures in e.g. bad weather or when blockage of the rails occur. An example of a predefined recovery rule could be a reduction of the frequency on all train lines in the case of heavy snow. A more detailed example is available in Groth et al. [2007], Section 3.3. Other recovery rules regard the turning of trains before their planned terminal, cancellation of trains etc.. Three such recovery rules are evaluated in a simulation in Hofman et al. [2006].

In the next sections we will discuss the recovery action seen from respective the network traffic controller, the crew dispatcher and the rolling stock dispatchers point of view.

4.2.1 Timetable adjustments at S-tog

We will in this section present timetable adjustments as it is conducted by BaneDanmark in the S-tog network. For a recent and comprehensive discussion of real time train routing and timetable dispatching in general, see the thesis by D'Ariano [2008].

As described in Groth et al. [2007] recovery is shared among three dispatching groups counting timetable adjustments, rolling stock and crew rescheduling. At S-tog the network traffic controller has the main responsibility for recovery of the timetable in case of disruptions. The real time train routing through stations and different sections of the network are conducted by local traffic controllers. Train routes are normally set automatically but can if necessary be set manually from the traffic control center.

At S-tog the local traffic controllers and the network controller are located at the same traffic control center. Information flow is often oral and therefore immediate. The network traffic controller applies different recovery strategies when dispatching the timetable departures. We refer to Groth et al. [2007] for

a comprehensive discussion of timetable adjustments.

Given that the changes to the timetable caused by either recovery strategies or the initial disruption directly are significant, the rolling stock and crew plan will be disturbed. Recovery actions that may cause disruption in the rolling stock or crew schedules are e.g. cancellation of train departures that causes a misplacement of the rolling stock and crew. For the rolling stock this means that at some point it is either not available or it is located wrongly for covering its scheduled train tasks.

4.2.2 Crew dispatching

The crew dispatcher has the responsibility of assigning train drivers to train tasks which in the operation is not yet covered. Also, they must update the personnel dispatching system. Today no decision support is available and surveillance systems must be updated manually. Ongoing projects in S-tog will ensure that the processes are automated to provide decision support. The train driver recovery problem (TDRP) is presented in Rezanova and Ryan [2006]. The solution approach described in the paper will be implemented at S-tog.

4.2.3 Rolling stock dispatching

The rolling stock dispatcher has the responsibility of assigning train units to train tasks not yet covered by rolling stock. He also finalizes the rolling stock depot plans according to short-term changes, and dispatches the Copenhagen rolling stock depot.

The rolling stock dispatcher monitors the condition of the running rolling stock and communicates with the workshop concerning defects on train units. He classifies occurring defects using a severity index indicating how soon a train unit must return to the workshop given the defect in concern.

Finally, the rolling stock dispatcher has the responsibility of deriving the reinsertion schedule as requested by the network traffic controller, see Groth et al. [2006]. This is in the specific incident where a train line has been cancelled. The reinsertion commences when the network traffic controller evaluates the operation to be stable.

The rolling stock control system does not update automatically according to real time changes. The rolling stock dispatcher therefore does the updating

manually. The rolling stock dispatcher maintains the knowledge of the locations of train units in the network from communicating with the network and local traffic controllers, communicating with the shunting personnel at the depots and by being connected to the train drivers by radio contact. In the same way he surveys the condition of the train units.

When disruptions occur in the rolling stock schedule the result may be that for some train departures the train units planned to cover them are no longer available i.e. the departures are uncovered. Also, a disruption to the rolling stock plan can mean that train units, in the process of recovering the timetable departures, are re-assigned to train departures that will leave them at a "wrong" rolling stock depot at the end-of-day. This may cause an imbalance in the number of train units available for the start-up of the operation the following day.

In the process of recovery there are several objectives to be considered. A certain *service level* must be obtained meaning that an allocation of train units to train tasks must be done under the consideration of the expected passenger demand. The objective of service level conflicts with the objective of *efficiency*. One of the greatest expenses of operating rail is the driven unit kilometers. To be competitive the rail operator must minimize the number of train units allocated to each train task. A third objective is *robustness* of the rolling stock schedule. Robustness is high when the number of composition changes is low. This is likely to conflict with the objective of driving with the exact number of train units needed to cover expected passenger demand. A last objective concerns the *deviation* from the originally planned schedule. Considering constraints on train units versus train drivers no union rules apply to train units and the train units do not have a subjective opinion on which depot to be on at the end of recovery. However, if train unit are close to their maintenance limit, they may not be able to get to the workshop within the maximum time and kilometer allowed. If they then end up at a wrong depot at night, empty stock transport may be necessary to get the train to the workshop. Also, if there is not the planned amount of train units available at depots at the end-of-day, the disruption will continue into the following day.

4.2.3.1 Applying a rolling stock recovery system

In this section we briefly describe how an automatic decision support tool for rolling stock recovery could be embedded in the operation.

At some point in time where the disruption is sufficiently large or the end-of-day balance restrictions on depots cannot be met, the Rolling stock Recovery Prob-

lem (RSRP) is delimited, run and a solution is found. We say that *Rolling Stock Recovery* is invoked if the *start criterion* is fulfilled. It is evaluated whether the solution is applicable in the operation, i.e. if it minimizes the end-of-recovery balance differences and cover the expected passenger demand sufficiently. If the solution is applicable implementation begins. As the operation is running, further disruptions in the rolling stock schedule can occur during the implementation of the solution. At any time if the *start criterion* is fulfilled a new instance generation and solution process of the RSRP begins.

Often there will be only one train task to cover, that is, the disruption will be strongly limited. The choice of train unit(s) to use will be limited to identify one of the train units on the departure depot of the train task which are available for coupling and makes the best match with respect to maintenance requirements and end-of-day depot. A limited case of the RSRP formulation can be constructed consisting of only the one train task or train sequence and the train units being located on the departure depot of the train task. However, the limited choice of rolling stock units on the departure depot may not lead to a good solution. A low quality solution, however, is often better than no solution. The solution found will therefore likely be implemented in the operation, but may contribute to a disruption in the rolling stock schedule which must be solved at a later point during the operation.

4.3 Conclusion

We have presented the three different areas of dispatching with a specific emphasis on rolling stock dispatching and the different ways the areas of dispatching correlate.

As is evident most of the decisions made in dispatching today are ad hoc and to a large extent subjective. This again means that the quality of a recovery depends strongly on the skills, experience and preferences of the individual dispatcher. Especially the decisions of the network traffic controller has a great impact on the quality of the recovery as these decisions can in fact create disruptions in the rolling stock respectively the crew schedules. Also, as the three areas of timetable adjustments, crew re-scheduling and rolling stock re-scheduling are handled separately any solution is very likely sub-optimal. Automatic decision support systems will hence most likely improve quality of the operation.

Only little research has been done concerning disruption management in rail according to Groth et al. [2007]. This is likely one of the reasons why there is only little use of automatic decision support in practice. The decision support

tools and/or the basic formulations and solution approaches simply do not exist in practically applicable form.

Rolling stock recovery - discussing network structures

In this chapter we discuss different representations of the RSRP. We start by discussing different flow formulation approaches as the structure of the RSRP is similar to problems earlier formulated using a flow perspective. As the RSRP is a problem occurring in real time, it demands short computation times. We therefore study different formulations with the objective of reducing the size of the formulation where size is measured in number of arcs in the networks or number of variables and constraints in the mathematical formulations.

5.1 Problem description

As we described in Chapter 4 the main responsibility of the rolling stock dispatcher is to ensure the continued coverage of train departures with rolling stock. In the case of a disruption this process becomes complex and hard to handle manually i.e. finding train units to cover the train tasks without any computer-aided decision support is time demanding. Also, it is hard to maintain a general view of the location of all train units and their individual constraints regarding maintenance.

Given a disrupted rolling stock schedule and a set of train units. Find a feasible re-routing of the train units so that the train departures of the rolling stock schedule are covered sufficiently given their estimated passenger demand, and such that the train departures assigned to each train unit form a feasible train task route

Table 5.1: The Rolling Stock Recovery Problem

We say that a rolling stock plan is disrupted when either some set of train tasks in the plan are not covered sufficiently relative to the passenger demand of each train task or when one or more train units are assigned to train tasks in such a way that at some point a deadheading will be necessary to relocate the train units in the railway network. Relocation of train units is necessary if, for example, a train unit is defect and therefore not able to carry passengers or if too few train units are available on a rolling stock depot to cover future train departures.

This chapter discusses the feasibility problem of rolling stock recovery where the basic problem is formulated in Tab. 5.1.

5.1.1 A decomposed approach

The problem can in a natural way be divided into two of finding composition for each tasks and hereafter routing the specific train units. First, train unit types are assigned to train departures (train tasks). For each train task the train types will have some order indicating the composition. Capacity of train units on the depots is controlled after the arrival of each train task so that no more train units than what is available is used from the depots. It must also be controlled that no illegal composition changes are made. A composition change is legal when train units are removed from or coupled to the open end of the composition. Fig. 5.1 illustrates the open-end concept.

After finding the composition for each train task the individual train units must be routed. All train tasks must be covered by train units given the information on how many train units of each train type must be assigned. This information is given by the solution found in the previous phase of deciding compositions. The train tasks assigned to each train unit must form a feasible work plan. This implies that given two consecutive train tasks assigned to the same train unit the arrival depot of the first train task must be the same as the departure depot

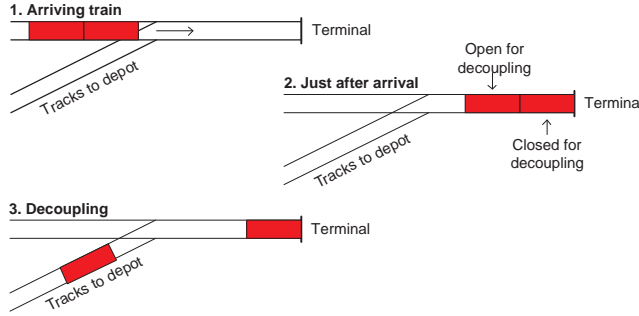


Figure 5.1: Illustrating the open and closed ends of a composition

of the second train task. Also, the departure time of the second train task must be later than the arrival time of the first train task.

The RSRP is split into two subproblems. This may limit the degrees of freedom in finding solutions. However, the complexity of the complete problem makes the split inevitable. Degrees of freedom are limited as binding the type and number train units used on each train task might invalidate some good solution with respect to the re-balancing of train units. For example, in practice a dispatcher may very well choose to put more or less train units on a train task than what is required according to forecasted demand.

5.1.2 Disruption time window

The extent of the disruption, (*the recovery scenario*) to be solved, is defined by a time span of the disruption and a set of train units.

The correlation of delays between train units assigned to train sequences of the same train line is very strong. That is, if one train sequence is significantly delayed, all the other train sequences on that train line will also be delayed. The recovery scenario is therefore defined as a set of one or several train lines affected. This means indirectly that a certain set of train sequences is involved in the recovery process and so are the train units assigned to them. In addition to the train units assigned to the train sequences the train units on the rolling stock depots on the route of the train lines involved are also included.

Typically a disruption is resolved by considering a time window within which it will be attempted to return to normal (planned) operation. The size of the time

window is critical to the computation time of any solution procedure, however, we do not wish that the time window limits the possibility of finding a good recovery solution.

There is no guarantee that a complete recovery solution can be found i.e. a solution where all train tasks are covered and all train units are routed to a preferred rolling stock depot. In that case there are two possible approaches.

In the first approach the solution found is used partially, meaning that train units are assigned to train task according to the solution found. The disruption is then redefined containing new train tasks and train units within a redefined recovery time window.

In the second approach the disruption scenario is expanded. Expansion can be made by considering an extended time window - thereby including more train tasks and train units in the disruption. The train tasks may be on the same or a different train line.

5.2 Discussion of network structures

In this section we will consider flow formulations of the RSRP.

In Ahuja et al. [2005b] a time-space formulation for the real-life locomotive-scheduling problem is presented. The problem resembles the RSRP. The main differences is that in the locomotive scheduling problem time is available for rearranging the locomotives on the depot stations in such a way that train units placed in the middle of a composition¹ can also be decoupled. This implies that the very complex composition constraints can be left out. Also, in the locomotive scheduling problem the objective is to decide the number of each type of locomotive to assign to each train task. This is a simplification of the RSRP where constraints are on each individual train unit. In the RSRP the individual train units each represent a commodity. Fig. 5.2 shows an extract from the time-space network presented in Ahuja et al. [2005b].

The network is composed by the train task, the depot and the connection arcs and of a set of source-to-depot/task, depot/task-to-sink arcs. A feasible locomotive schedule maintains the flow balance constraints for each node.

We will now present a time-space based multi commodity flow model for solving the RSRP. The model is based on the network presented in Ahuja et al. [2005b],

¹A composition is denoted a consist in Ahuja et al. [2005b].

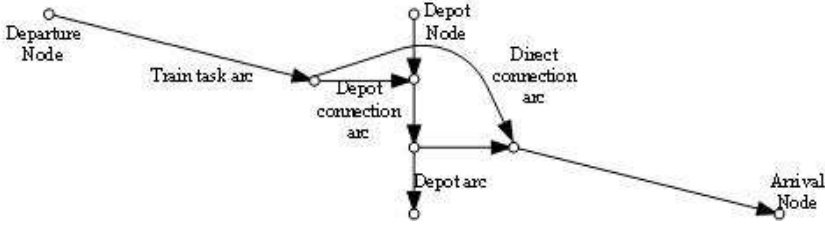


Figure 5.2: An extract of the network used for the real-life locomotive scheduling problem in Ahuja et al. [2005b]. The figure shows two tasks connected to each other and to a depot.

where the commodities in the model are the physical train units.

The nodes in the network are partitioned into four different sets: Event nodes, V_e , which represent either departures or arrivals of train tasks, ground nodes, V_δ , which are the depot nodes, and source and sink nodes, resp. V_{so} and V_{si} , which are not illustrated in Fig. 5.2. The joint set of nodes is referred to as V .

The arcs in the network are subdivided into 5 different sets: The task arcs, E_{tasks} , connect a departure event node and a arrival event node. The decoupling arcs, E_d , connect an arrival node with a ground node, and the coupling arcs, E_c , connect a ground node with a departure node. The direct connection arcs, E_γ , connect an arrival node with a departure node of the subsequent train task². Finally, there are the depot arcs, E_δ , which connect the ground nodes within each depot. The joint set of arcs is denoted E .

The sets I and J with indices i and j respectively denote nodes in the network. We say that $(i, j) \in E$ is an arc in the network, if there is a direct connection between i and j . K is the set of physical train units indexed by k , and D is the set of depots indexed by d .

The mathematical model is based on a set of binary decision variables, x_{ij}^k for all $i \in I, j \in J$ and $k \in K$.

$$x_{ij}^k = \begin{cases} 1 & \text{If train unit } k \text{ is assigned to the arc between } i \text{ and } j \\ 0 & \text{Otherwise} \end{cases}$$

²Recall that a subsequent train task is the direct successor task within the train sequence of the train task in question.

Model

$$Z = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} (Demand_{i,j} - Seats^k) \cdot x_{ij}^k + \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ij}^k \cdot x_{ij}^k \quad (5.1)$$

$$\sum_{\substack{j \in J, \\ (i,j) \in Out_i}} x_{ij}^k - \sum_{\substack{j \in J, \\ (j,i) \in In_i}} x_{ij}^k = \begin{cases} U_i^k & \forall i \in V_{so}, \quad \forall k \in K \\ 0 & \forall i \notin V_{so}, V_{si}, \quad \forall k \in K \\ U_i^k & \forall i \in V_{si}, \quad \forall k \in K \end{cases} \quad (5.2)$$

$$MinL \leq \sum_{k \in K} Length_k \cdot x_{ij}^k \leq MaxL, \quad \forall (i,j) \in E_{tasks} \quad (5.3)$$

$$\sum_{\substack{i \in I, \\ j \in J}} KM_{ij} \cdot x_{ij}^k \leq KmLimit^k, \quad \forall k \in K \quad (5.4)$$

$$\sum_{k \in K} x_{ij}^k \leq Capacity_{ij}, \quad \forall (i,j) \in E \quad (5.5)$$

The objective, 5.1, is to minimize the total sum of standing passengers plus the total cost, c_{ij}^k , of covering edges where a cost is assigned to each edge. The two terms are conflicting in the sense that while you want to ensure the coverage of the passenger demand you still want to reduce the flow through the network to ensure the efficiency of the solution, see 4.2.3. Eq. 5.2 ensures feasible flow. The parameter U_i^k gives for all source nodes the number of units available of train type k . The lower and upper limits of train units on train tasks are included in Eq. 5.3. The parameters $MinL$ and $MaxL$ gives respectively the minimum and maximum length of a train. In Eq. 5.4 the sum of driven kilometers for each train unit is limited by an input parameter $KmLimit_k$. Finally, in Eq. 5.5 the capacity on each arc is limited by an input parameter $Capacity_{ij}$, which varies for the individual arc, $(i,j) \in E$. For $(i,j) \in E_\delta$ the capacity parameter equals the maximum number of train units that can be located in the depot at a time. For $(i,j) \in E_{tasks}$ the capacity parameter is maximum number of train units on a train.

The problem with the time space network presented in Ahuja et al. [2005b] is that composition changes cannot be taken into consideration.

For later comparison, the maximum size of the problem is $3.3 * 10^8$ variables. Most of these will be zero as the network only contains approximately 3000

arcs. Given a maximum of 100 train units available in the recovery scenario potentially 300,000 variables can assume a value differing from zero. In reality this number is considerably lower. There are approximately 200,000 constraints in the model.

We will in the following consider different network formulations for the RSRP. First, we will consider an integrated approach for addressing the problem.

5.2.1 Extended network

To be able to create feasible compositions the time space network must be expanded. One way to do this is to add extra nodes and arcs representing the positions of the multiple train units. The nodes at each departure or arrival represent the positions in the departing or arriving train respectively. The arcs connecting the train task events of departure and arrival are connected to nodes in such a way that the position of a train unit is changed if the train is open for decoupling in its other end at the next event e.g. when it arrives at a terminal depot. In Fig. 5.3 a reduced version of the network is shown having only two train tasks connected to a depot.

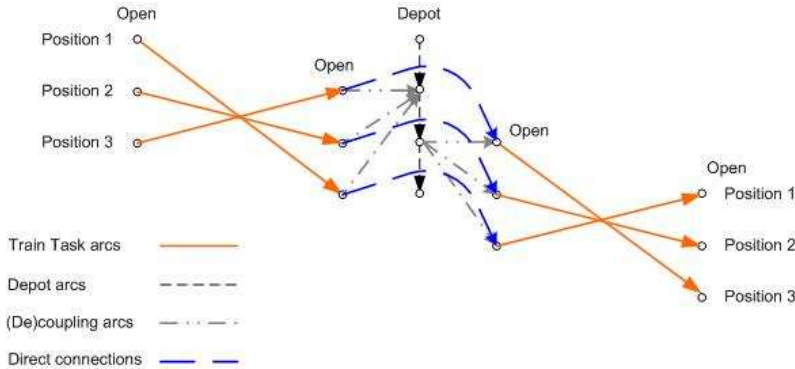


Figure 5.3: A time space network expanded to handle composition changes. Three arcs represent each train task. Also, decoupling and coupling are represented by three arcs each according to the decoupling/coupling from/to each position in the composition.

In Fig. 5.4 an example of a flow in the network is illustrated. Two train tasks are illustrated, $task_1$ and $task_2$. The first train task, $task_1$ is covered by two train units, $unit_1$ and $unit_2$ whereas $task_2$ is only covered by $unit_1$ that is $unit_2$ is decoupled at the arrival depot of $task_1$. $unit_1$ is in the open end at

the departure depot of $task_1$, and as the open end shifts between the departure and arrival depot of $task_1$, $unit_2$ can feasibly be decoupled. It is decoupled and continues along the ground arcs.

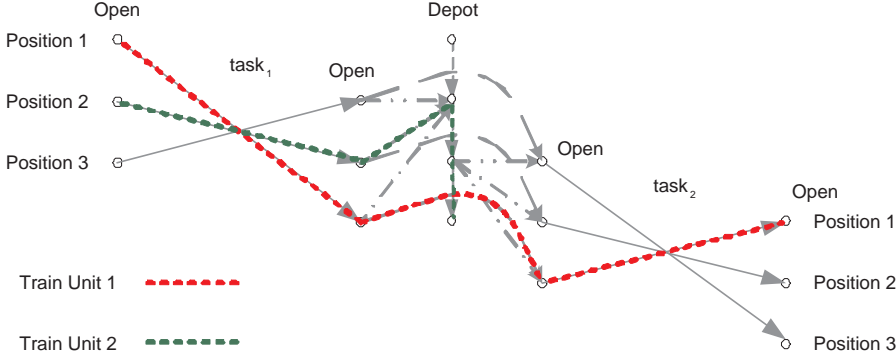


Figure 5.4: Example of flow of two train units through a composition change flow network

Decoupling can only occur from the open end of the train i.e. from the top arrival event node. Likewise, coupling can only occur to the top departure event node of the subsequent task. Direct connection arcs are added for those train units that are not coupled/decoupled between two consecutive tasks.

When the open end of a train task does not change from one station to the next, the event arcs do not cross. That is, the top departure event node is connected to the top arrival event node.

The nodes and arcs of the network are represented as in the basic flow formulation, see Fig. 5.2. The event nodes and arcs, and the decoupling and coupling arcs are however divided in several groups each with respect to the different *positions* within the composition on task. There are three event nodes for each departure referring to respectively *position* 1, 2 and 3 in the train composition. We denote them $V_{\epsilon(d)_1}$, $V_{\epsilon(d)_2}$ and $V_{\epsilon(d)_3}$. Equally, there are three nodes for each arrival, $V_{\epsilon(a)_1}$, $V_{\epsilon(a)_2}$ and $V_{\epsilon(a)_3}$.

The event arcs are grouped into three sets, E_{ϵ_1} , E_{ϵ_2} and E_{ϵ_3} . When the open end is in the same end of the train on the two stations of an event there is a node from *position* 1 on the departure station to *position* 1 on the arrival station etc. When the open end shifts from the departure to the arrival station *position* 1 is connected to *position* 3 and *position* 3 is connected to *position* 1. E_{ϵ_1} are the set of arcs emanating from *position* 1, E_{ϵ_2} are the set of arcs emanating from *position* 2 and E_{ϵ_3} are the set of arcs emanating from *position* 3.

Also, decoupling respectively coupling arcs are now represented by three arcs for the flow between positions and the ground node in each composition. The sets $E_{d_{p_1}}$, $E_{d_{p_2}}$ and $E_{d_{p_3}}$ represent the decoupling arcs from respectively *position* 1, 2 and 3 to the ground node. The sets of arcs $E_{c_{p_1}}$, $E_{c_{p_2}}$ and $E_{c_{p_3}}$ represent the coupling arcs from the ground node to *position* 1, 2 and 3 respectively.

The direct connection arcs are also divided in three sets, E_{γ_1} , E_{γ_2} and E_{γ_3} .

Of course adding all these arcs to handle the composition increases the size of the problem considerably. The size of the problem potentially exceeds 1 billion variables in total. Most of these can be fixed at zero as there are less than 6,000 arcs in the network in Fig. 5.3. Given that maximum 100 train units will be available in the recovery scenario, potentially 600,000 variables can assume values differing from zero. In reality the number will be lower. The number of constraints is increased to more than 360,000.

5.3 A decomposed approach

We will now consider the decomposed approach for the RSRP of firstly finding compositions for each train task and secondly routing the train units in paths covering the train tasks.

5.3.1 Finding compositions

The first of the subproblems in the RSRP is the problem of allocating train unit types to train task so that train tasks are covered sufficiently with respect to passenger demand, balances at rolling stock depots are kept and composition changes (coupling and decoupling of train unit between train tasks) are feasible.

Alferi et al. [2006] present a transition network formulation for the first of the subproblems of the decomposed approach. Nodes represent feasible compositions for train tasks and the arcs represent the possible composition changes between tasks. The compositions available for each train task is limited according to the estimated passenger demand. For each composition on a specific train task there is a set representing the possible compositions for the successor train task. The model presented in Fioole et al. [2006] is based on the transition network structure. We will refer to the model as the *composition* formulation. It can easily be adapted to the S-tog problem. It is sufficient to use the model without extensions for splits and merges. The main decision variables are the

binary variables being one if train task t has composition p and the successor of t , $\nu(t)$, has the composition p' . We have calculated the size of a real time example of this formulation on a S-tog instance where the compositions are allowed to be either up to two train units or up to three train units. The problem scenario covers a two hour time window in the morning peak hours including all train lines covering the central section. The number of variables in the formulation having maximum two unit compositions is approximately 27,000. In the formulation having maximum three units per composition it is approximately 41,000. The number of constraints are respectively less than 12,000 and 16,000.

We have reduced the problem size further by formulating the composition problem in an assignment model assigning train unit types to the specific positions in the compositions for each train task. We will refer to it as the *position* formulation. The main decision variables of the model are binary and one if train type m is assigned to train task t in position p . The formulation could be tractable for S-tog since the maximum number of train units that can be assigned to any composition is low. However, the composition changes must in the *position* formulation be handled in the constraints of the model. The detailed description of the *position* model is given in Groth et al. [2008].

5.3.2 Train unit routing network

We now know the composition of each train task meaning that we know the demand for train units of each train type for the train tasks. Based on this we now need to route the physical train units.

The time space network in Fig. 5.2 can directly represent a network for routing train units. A reasonable maximum number of train tasks in a recovery scenario is approximately 400. In the time space network the addition of one train task to the recovery scenario means the addition of 6 arcs i.e. 400 train tasks means an approximate number of arcs of 2,400. If 100 train units are available for the recovery process this means approximately 240,000 variables representing the arcs alone. Potentially there are even more variables in the model. We aim at reducing this number of variables and do so by modifying the network structure.

In the time space formulation two event nodes of departure and arrival represents each train task. We can be sure that if a train unit covers a departure node it will also cover the following arrival node, we therefore join the nodes into one. That is, T is the set of train tasks and $|T|$ is the number of train tasks. The $2 \cdot |N|$ nodes representing train task departures and arrivals are aggregated to a set of $|T|$ nodes representing all train tasks. In Fig. 5.5 an example of the network is illustrated.

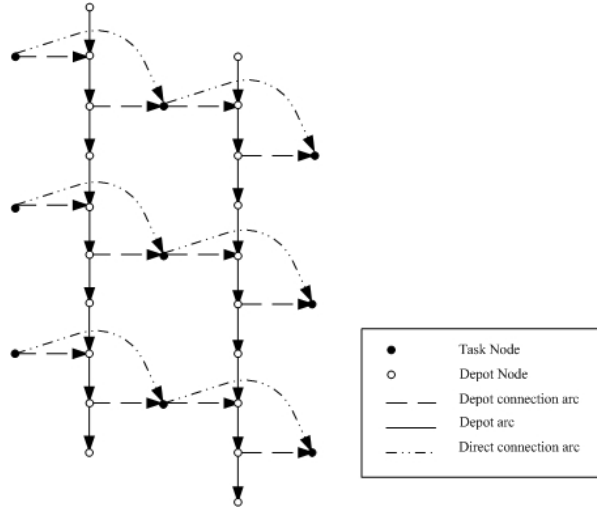


Figure 5.5: A connection network including depot nodes where departure and arrival events are gathered in one node.

The aggregated network in Fig. 5.5 is a modified connection network. For a description of a standard connection network, see e.g. Clausen et al. [2005]. We call it modified as the network still includes the depot nodes. The network has approximately 2,000 arcs. Given maximum 100 train units in a recovery scenario there are potentially 200,000 variables in this model that can assume a value differing from zero. As earlier, in reality this number will be lower.

If we leave out the depot nodes, the connection network is as in Fig. 5.6. The train tasks are now connected directly to the possible successor train tasks i.e. connecting train tasks are those departing from the arrival station of the train task after the arrival time of it.

Let us include the time perspective in the nodes by adding a time line and rearrange the train task so that the train tasks belonging to the same train sequence lie on the same horizontal line in the figure. Also, let the nodes illustrate their respective time span, see Fig. 5.7. The network is now presented in a Gantt chart structure. If we consider a basic flow formulation based on this network modification and assume that a maximum recovery scenario has around 400 tasks there are potentially $|T| \cdot |T| \simeq 240,000$ binary variables representing flow on arcs. Given the 100 train units this means that a flow formulation would have up to 2.5 mil. variables representing flow alone. Again, in practice the number of arcs in the network will be lower.

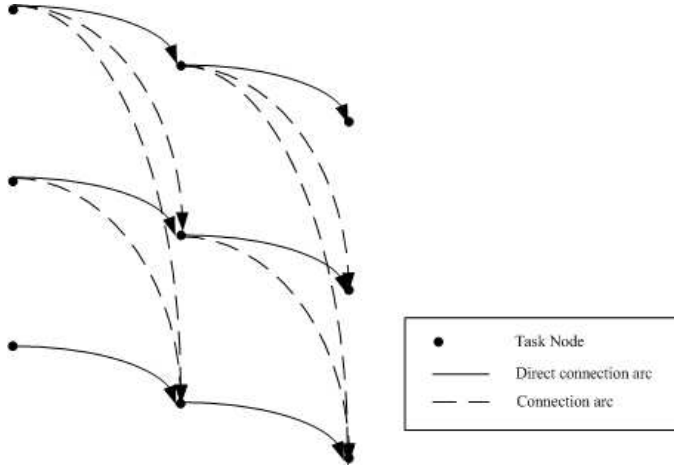


Figure 5.6: A connection network on the train routing problem.

The size of the graph is limited by the span of the disruption time window defined prior to the construction of the graph. A practical example from S-tog containing all train tasks on all lines in a three hour time span contains fewer than 22000 arcs.

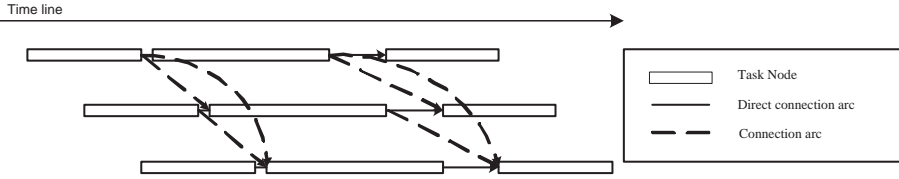


Figure 5.7: A connection network on the train routing problem with a train sequence alignment and including the time perspective in the node time span.

We call the network illustrated in Fig. 5.6 the graph of the train routing problem. Again, the graph G is directed where V is the set of nodes and E is the set of edges. A node $v \in V$ represents a train task departing at departure time τ_{Depart} from station δ_{Depart} and arriving at arrival time $\tau_{Arrival}$ at station $\delta_{Arrival}$. An edge between to nodes v and w represents that the train task of node w can be carried out after v i.e. the two nodes can be covered by the same path. A path (or a train route) in the network corresponds to a work plan for a specific train unit.

Each train task equivalent to a node in the graph is assigned a length in kilometers relative to the distance between the departure and arrival station of the train task. Maintenance constraints for a train unit are kept, if the sum of kilometers for the train tasks assigned to the train unit in the recovery phase plus the sum of kilometers the train unit has driven prior to the recovery and will drive after the recovery is less than the kilometer limit on driven kilometers between two maintenance checks.

Furthermore there are time maintenance constraints. These measure the actual time since last maintenance check. There is no difference in handling kilometer respectively time maintenance constraints.

The network has a set of sources and a set of sinks. The sources represent the train units available for that specific network. For each train unit the number of driven km and its seat capacity is given. The sinks represent the depots in the recovery scenario that each train unit potentially can end on for a given path.

We have reduced the problem size further by representing the routing problem by an assignment model instead of a traditional multi-commodity flow model. The main decision variables in the assignment model are the binary variables, Q_t^k . The mathematical model is presented in Groth et al. [2008], Appendix E, however, we will go into some details concerning the construction of feasible flow.

In the assignment model feasible flow solutions are constructed indirectly by adding constraints to control that for each train task assigned to a train unit, except for source and sink, there are always at least one predecessor and at least one successor assigned also. Each train unit is assigned to at most one source and one sink. At the same time, a set of constraints ensures that the flow found for a train unit is unambiguous, that is, a train unit must cover no more than one train task at a time. Train tasks are considered to be at the same time when they have a non-empty intersection. We call this concept "time-parallelism" and we avoid that a train unit covers time-parallel train tasks by the constraints stating that if a train unit, u , is assigned to a train task t no other train tasks, which are time-parallel to t can be assigned to u .

The flow is unambiguous for a train unit if for each train task in its path only one true predecessor respectively successor³ is assigned. The flow from node to node in the network is not controlled explicitly in the assignment model. Therefore we must ensure that at least one in the set of predecessors respectively successor of a train task is covered.

³A true predecessor/successor to a train task is the immediate predecessor/successor train task on the same train sequence

Considering a longer time perspective there will exist pairs of train tasks, t_1 and t_2 , which are not time-parallel and both successors to a train task, t . Given the constraints in their present structure, a single train unit can be assigned to all three train tasks. This can imply that the path of the train unit is split into two, see Fig. 5.8. As the two orange train tasks are not time parallel, they can be assigned to the same train unit. However, the reduced size of the time period considered in a recovery situation implies that no train tasks exists that are not time parallel and can be covered by the same train unit. Therefore, split paths can occur.

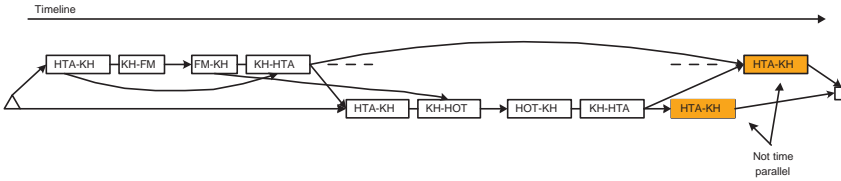


Figure 5.8: When considering a wide time perspective assignments can be made that result in a split path. If the two orange tasks are assigned to the same train unit the path is split in two.

5.4 Conclusion

In this chapter we have discussed several ways to construct the network/-s of either the integrated or decomposed RSRP. The purpose has been to get a clear indication of whether to use a decomposed or integrated approach for the RSRP. The models presented in this chapter indicate that the flow models and integrated models will be too extensive for achieving attractable solution times. The final choice of models decided to be implemented, are described in Groth et al. [2008] and chosen mainly because of their relatively limited sizes with respect to number of decision variables. A sufficiently low computation time is crucial to the applicability of an RSRP solution approach. The number of variables necessary to represent the flow formulation presented in this chapter indicate that computation time will be too large. In Groth et al. [2008] we describe a decomposed approach for the RSRP with models of reduced size. We will discuss the paper in more detail in Chap. 6.

CHAPTER 6

Discussion of thesis papers

In this chapter we sum up and discuss the four papers included in the thesis. We discuss the papers from the perspective of model complexity, time, and applicability. Specifically we concentrate on the three papers contributing with simulation or optimization models. The survey paper in App. B will be complementary to the two Chapters 3 and 4 of the thesis.

The three papers will be presented in order of increasing problem complexity. Paper C presents a simulation model. Paper D an optimization model for reinserting a cancelled train line is presented. The model is implemented in Gams and solved by CPLEX. The model for rolling stock recovery presented in the paper E is an advanced optimization model with features beyond simple reinsertion. It has been implemented in C# using ILOG concert technology to call CPLEX for each of the three sub models.

The simulation model in C considers the operation at a macroscopic level based on the level of detail in the public timetable. The optimization model in D is a little more detailed also considering the trains on each train line and the identification numbers on the departures of the trains to ensure that each train on a train line is only inserted once. Again the model with the highest complexity is presented in the paper E where all details of the timetable are included.

The simulation model presented in C is used in the strategic and tactic planning

phases as an evaluation tool prior to the operation. The two optimization models in papers D and E are both meant for dispatching and recovery in real time.

Summing up over the different perspectives we see that the models increase in complexity as we approach the real time operation.

6.1 Disruption management in passenger railway transportation

The paper, co-authored by people from the Dutch railway company NS Reizigers (NSR), gives a comprehensive description of disruption management in passenger railway transportation. The different actors in the disruption management process are presented: Local and global network controllers, who are representatives of the infrastructure manager, and local and global operations controllers, who are representatives of the railway operators operating the network. The organizational challenges of being several interested parties in a decision process are discussed. Also, the three subproblems related to disruption management: Network traffic control, rolling stock and crew re-scheduling are presented and their processes are described. The problems are described in a context giving an indication on how they are interdependent and what other characteristics affect the operation. The different problems are exemplified by practical examples from respectively S-tog and NSR.

The main purpose of the paper is to give an introduction to disruption management in a railway operation context. One conclusion is that only little work have been made to refine existing planning techniques with rolling stock and crew to be capable of performing good results and form the basis of decision support for the real-time operation. Also, only few research studies have presented results from new models and methods presenting and solving the real time problems.

It is important that there is a clear understanding of the partition into three subproblems of the operation. It is a common organizational structure that the responsibilities are divided among infrastructure manager and railway operator such that the first is responsible for network traffic control and the latter is responsible for the resources of rolling stock and crew. Furthermore, the railway operator commonly divides the responsibility of dispatching the crew and rolling stock among two separate dispatching units. There are, however, clear interdependencies between especially network traffic control and rolling stock re-scheduling on one side and network traffic control and crew re-scheduling on the other.

It is a challenge to formulate decision support tools that are able to function across two organizations. Also, there must be a close communication between the problem areas. Close communication is best supported by physically placing the dispatchers of the different areas in the same room i.e. centralized dispatching and control center.

6.1.1 Contribution of paper

The paper is the first overview of disruption management in the passenger railway industry. It describes the interrelations between infrastructure owner and railway operator and the different objectives of the two. The main problems of real-time dispatching are outlined.

6.2 Robustness and recovery in train scheduling - a simulation study from DSB S-tog a/s

In this article we present a simulation model used for two robustness studies. The model is a macroscopic simulation model addressing the timetable at a time perspective level meaning that the entities driving the simulation model is departure events which are disturbed by stochastic incidents. The probability of these is based on a empirically derived distribution function.

The first study is on robustness of different basic structures of timetables. Different timetables are constructed and tested for their ability to absorb delays. We define a timetable as robust, if it is able to absorb minor delays i.e. if the number of knock-on-delays is low. Timetables with significantly different structures are used in the analysis.

The second study investigates three different recovery strategies used at S-tog. S-tog has a set of preplanned recovery strategies used in the operation when disruption occurs. The recovery strategies are used in the operation based on the subjective choice of the network traffic controller. The network traffic controller evaluates the severity of a disruption and then makes a choice of how to recover. The recovery strategies evaluated with the simulation model are first *cancellation of train lines*, second *turning trains around early* and third *insertion of on-time train on the central station*. In the first recovery strategy trains on an entire line is taken out of operation by shunting them to depots when they reach these along their route. The second recovery strategy is that if a train is delayed on its route going from terminal A over station B to terminal

C, the train is turned early on e.g. station B, thereby shortening the route of the train and saving time. In the third recovery strategy an on-time train is inserted at the central station instead of a train having been delayed earlier on the route. When the delayed train arrives at the central station it is taken out on the central station depot.

The results of both studies are applicable in practice. From a strategic planning perspective the simulation suits the purpose of evaluating new timetable structures before effort is put into planning the timetable in detail, and more importantly knowledge is obtained on how robust the timetable is in practice relative to e.g. the present timetable. The second study reveals important knowledge on the disruption sizes relative to the choice of recovery strategy.

As mentioned, the simulation model presented in the article is macroscopic. This is opposed to standard rail simulation models which are microscopic. We believe a macroscopic model is sufficient for evaluation of robustness. An interesting further study could be to use a microscopic simulation model on the same timetables as our model. One of the challenges linked with applying a microscopic model is that the microscopic level requires many details before a simulation can be performed. This implies that ideas for new timetables must go through a major part of the detailed planning process before we can simulate it. In a macroscopic model the overall timetable structure is sufficient to use the model for evaluation.

A further study could be to construct a microscopic model with the same precondition as the macroscopic model and subsequently comparing the conclusions of the two models.

6.2.1 Contribution of paper

The paper contributes with the interesting aspect of developing a macroscopic simulation model for evaluating the robustness of railway timetables and for testing the effect on robustness of three recovery strategies.

The paper presents several interesting results. First, the results show that there is an upper limit on how much buffer time in the timetable leads to an improved punctuality. Second, small delays are not in themselves affecting the punctuality of the operation, however, when small delays occur in an already disturbed timetable the total disruption will increase. Third, when disruptions are large the recovery strategy applied must give maximum increase in headways.

6.3 Optimal reinsertion of cancelled train line

In this paper the practical problem of reinserting a cancelled train line is assessed and a mixed integer programming model is presented. Recall that a train line consist of a set of trains forming a circuit where the trains run with a constant frequency.

Train lines are cancelled when the disruptions in the network are sufficiently large according to the network traffic controller. The train lines cancelled have a route intercepting the disrupted section in the network.

When the operation is recovered and no more running trains are delayed, the cancelled train lines are reinserted. Today, reinsertion is carried out for one train line at a time. This is also a basic assumption in the model presented in the paper.

A mixed integer programming model was formulated using the basic knowledge that a train line consists of a set of trains driving in a periodic circuit. We refer to the model as the *Reinsertion* model.

Each train in the train line has at the point of cancellation been removed from the network at some depot. When reinserting the train line each train must be reinserted exactly once.

Each train is represented by train tasks, which again represent the timetable. Each train is covered by a composition consisting of one or more train units. The composition necessary to cover the expected passenger demand at time of cancellation is not necessarily the same as at the time of reinsertion. Also, the train units forming the composition at the time of cancellation does not necessarily need to be inserted at the same train at the time of reinsertion. Therefore, a train that has been taken out of the operation on some specific depot at the time of cancellation will not necessarily be reinserted at the same depot. This decision is left to the rolling stock dispatcher as a complementary decision to the one regarding which train should be reinserted from which depot. Of course, each train on a train line will be reinserted from some rolling stock depot located on the route of the train line.

The *Reinsertion* model is based on the reinsertion problem of S-tog. It can be used in all railway networks which has a certain set of properties. First, the timetable must be fully periodic. This means that we must be sure that all trains run with the same frequency. For example, every train bringing out a train driver must arrive at the same minute counting from the beginning of each frequency interval. Another example relevant to the *Reinsertion* problem

is that every reinserted train must depart at the same minute counting from the beginning of each frequency interval. Second, the circuit of a train line must not be too long. If a circuit is long it will not be a feasible recovery action to cancel all the trains covering it as this will imply a too high deterioration of the service quality offered to passengers. Third, S-tog apply recovery by cancellation of train lines because of the dense structure of the railway network crossing the central section of Copenhagen. In this case the cancellation of a train line creates sufficient buffer times in the timetable for the remaining lines to recover to the scheduled operation. The two first properties mentioned are necessary for the *Reinsertion* model to be applicable. The third property is an argument to why S-tog apply cancellation and hereafter reinsertion as a recovery strategy. There may be other properties in railway networks also making cancellation/reinsertion a tractable recovery strategy.

6.3.1 Contribution of paper

The paper presents a model which is the basis for a decision support tool for the reinsertion of train lines. The tool is fully applicable in the operation. The system is in use today and has since its original version been updated to fit the new timetable. It supplies optimal reinsertion schedules within a computational time close to zero.

The *Reinsertion* model and its implementation in the operation illustrates that the distance from mathematical formulation to practical tool can be short. The model is applicable for all railway operators who drive with a fully periodic timetable and use the recovery strategy of cancelling entire train lines.

6.4 Rolling stock recovery problem

In this paper we present a decomposed model for solving the RSRP. Different structures for the network of the RSRP was discussed in Chapter 5. The decomposed approach partitions the problem into three sub problems which are solved iteratively, see Fig. 6.1.

The first model, the *Position* model, finds the compositions for train tasks. It assigns train types to the individual positions of the compositions of train tasks according to the expected passenger demand. The assignment of train types to positions considers feasibility of composition changes and the difference in capacities on depots.

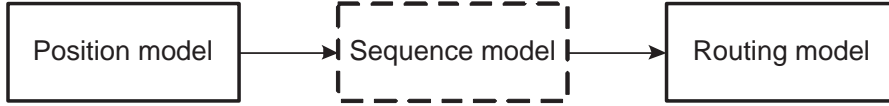


Figure 6.1: The decomposed approach for the RSRP. The dotted line illustrate that the Sequence model is optional in the solution process.

Secondly, the individual train units are assigned to train tasks so that they form feasible routes of train tasks. The two last models jointly assign physical train units to train tasks in a two-step approach.

The first step is an aggregated model, the *Sequence* model. Here, the train tasks are considered in sets of train sequences. At most one train unit is assigned to each train sequence. Therefore, no consideration of composition changes is necessary. For each train sequence a preferences is given to a train unit type, if the train type has been assigned to all train tasks of the train sequence in the *Position* model solution. The model maximizes the sum of preferred train units assigned to train sequences.

The second model assigning physical train units to train tasks is called the *Routing* model. It is an assignment model, which indirectly ensures feasible flow in the network¹ by adding a set of constraints which limits the number of train tasks assigned to a train unit at the same time to 1. We call this set of constraints parallelism constraints. The parallelism constraints and other sets of constraints of the *Routing* model are discussed briefly in Section 5.3.2 and the complete mathematical model is presented in Groth et al. [2008]. Additionally a set of constraints locking variables according to the solution found with the *Sequence* model are included. To control the solution quality a set of binary variables are introduced in the model being one if a train unit k is assigned to a train task, t , and its subsequent train task, $\nu(t)$.

The *Sequence* model is included to reduce the overall computation time of finding sufficiently good solutions. Technically it can be left out in the solution process, as it locks some of the variables at an aggregated level. If the *Sequence* model is left out of solution process, the computation time of the *Routing* model increases dramatically.

¹The network is presented in Chapter 5.3.2

6.4.1 Contribution of paper

The weakest link in the solution approach is the first problem of assigning train unit types to train tasks, the *Position* model. When the problem instance is relatively limited the computation time is low, however, the computation time increases as the problem size increases. We suggest that the recovery instances treated are strictly limited to only containing trains directly involved in the rolling stock disruption. If no satisfactory solution can be found the problem instance is iteratively expanded by a minimum of other train units or by a limited expansion of the recovery time window. At each iteration a solution is sought. This procedure will ensure that the size of the problem instance is at all times kept at a minimum.

When the *Sequence* model is included as a step of the solution approach the computation time of the *Routing* model is in general low and the approach becomes practically tractable. If a poor solution have been found in the *Position* problem the *Sequence* model does not necessarily lock many variables. This implies that there is more free variables in the *Routing* problem and the computation time will increase. In practice it will according to our test results only happen in few instances.

The practical applicability of the solution approach depends on the solution time. Based on knowledge of the rolling stock dispatchers work process, an estimate of the upper limit on solution time is around 5 minutes. The *Position* model solves within 70 seconds for problem instances with up to approximately 110 train tasks. The *Routing* model solves problem instances of up to 90 train tasks within 1 second.

The paper presents one of the first approaches for solving the RSRP. It is applicable in practice for instances up to at least 90 train tasks regarding solution time. The approach considers the complete recovery process of finding compositions for each train task and hereafter routing the physical train units.

A heuristic approach for the RSRP

In this chapter we will consider a heuristic approach for solving the integrated RSRP. That is, compositions for train tasks are found while building lines of work for train units. The heuristic approach is a work in progress.

We base our heuristic on the network illustrated in Fig. 5.7. The heuristic finds feasible train unit routes for each train unit in the network. The train unit routes must be feasible with respect to maintenance requirements. Also, each train task must be assigned train units according to its expected passenger demand. Finally, as more than one train unit can cover each train task we must ensure that train routes are feasible with respect to composition changes.

7.1 A construction heuristic

Initially, a feasible solution is constructed in a two-step construction heuristic. First, the train sequences included in the disruption are covered by at most one train unit each. In this process it will not be necessary to take into account the composition changes. After the first step a subset of train tasks will not have had enough train units assigned to it to cover its expected passenger demand.

In the second step of the construction heuristic we use a shortest path algorithm for covering the remaining train tasks which have not had sufficiently many train units assigned.

7.1.1 First step in the construction heuristic

The first step of the construction heuristic is based on knowledge of how a good rolling stock schedule is structured. The main characteristic of a good rolling stock schedule is that the train units in general follow one train sequence.

We will in this step consider each train sequence separately in a given order, i.e., train tasks belonging to the same train sequence will first be considered as a set. We will try to construct a fully or partial cover of each train sequence. A train sequence is fully covered when a train unit can be assigned feasibly to the joint set of train tasks on the train sequence. In a partial cover a set of adjacent train tasks on the train sequence is covered.

For each train sequence the heuristic searches for a train unit from the departure depot of the first task that constitute a feasible cover, that is, start depot of the train unit is identical to the departure depot of the first train task in the train sequence. The distances in kilometers of the assigned train tasks must not exceed the maintenance kilometer limit. The train unit must end on its desired end depot. Finally, a preference is given as input for train units of a train type where the number of seats available makes the overall best possible match between the average expected passenger demand of the train tasks and the seat capacity of the train type.

If no train unit can be found satisfying the constraints of maintenance and desired end depot, the set of train tasks on the train sequence is reduced and a new search for a feasible train unit assignment is made. That is, if no train units are available at the departure depot, the first task is left uncovered and the next is considered and so forth. The process continues until an assignment has been made or there are no more train tasks to consider in the train sequence.

The described process is carried out for all train sequences, see Algorithm 1. The function *ExistsCover*(ψ) returns a train unit which is a feasible cover of the train sequence ψ , if such one exists. *MakeCover*(ψ, v) makes a cover of train sequence ψ with train unit v . Finally, the function *RemoveFirstTask*(ψ) removes the first train task from a train sequence.

Algorithm 1 Construction heuristic, step 1

```

for  $\psi = 1$  to  $|\Psi|$  do
  if  $Length(\Psi) > 0$  then
     $v \leftarrow ExistsCover(\psi)$ 
    if  $v \neq null$  then
       $MakeCover(\psi, v)$ 
      STOP
    else
       $\psi \leftarrow RemoveFirstTask(\psi)$ 
    end if
  else
     $\psi \leftarrow \psi + 1$ 
  end if
end for

```

7.1.1.1 Order of execution

The quality of the initial solution depends heavily on the order in which the train sequences are handled. Train Sequences will be considered in order with respect to lines. Each train sequence has some start depot. The sequences will be handled for each start depot separately. Within the start depots the sequences will be handled in order with respect to the departure time of the first train tasks of the sequences, i.e., the earliest departure handled first.

7.1.2 Second step in the construction heuristic

As indicated, after the initial run of Algorithm 1, there will most likely be a subset of the nodes in the network where the expected passenger demand is not covered with sufficiently many train units. Hence, a second heuristic is used to cover the insufficiently covered nodes. The second step in the construction heuristic repeatedly constructs resource constrained shortest paths (RCSP) for each of the train units not used in Algorithm 1. A more general description of shortest path theory can be found in Ahuja et al. [1993].

We have constructed an algorithm for this based on the algorithm presented in Boland et al. [2006]. The paper deals with the Resource Constrained Shortest Path Problem (RCSP) which is adaptable to our problem. The paper presents a two-phased pre-processing procedure and a label-setting algorithm for finding the resource constrained shortest paths. First, the pre-processing algorithm reduces the number of nodes and arcs in graph following the procedure presented

in Aneja et al. [1983]. In the second step the all-pairs shortest path problem is solved. In the process any dominated paths, with respect to all resources, are eliminated. The paths leading to a node are indicated by labels attached to the node. Removal of paths by domination means that for two paths, ρ_1 and ρ_2 , leading to a node, if the resource usage and cost of ρ_1 are lower than that of ρ_2 , we say that ρ_1 dominates ρ_2 and therefore ρ_2 can be eliminated.

7.1.3 The label-setting algorithm

In this section we present details on how the label-setting algorithm presented in Boland et al. [2006] can be used on the RSRP.

The main objective is to cover all train tasks according to demand and to position the train units feasibly with respect to the further operation. The solution will consist of train unit paths covering the train tasks. As we want to minimize the number of train sequences for each train unit a cost is imposed on each change. Also, a penalty is added when adding a node which is already covered sufficiently with respect to expected passenger demand.

When constructing paths in the RSRP network these are limited in length by the maintenance constraints and by the number of composition changes. The maintenance constraints are quantified by the distance that the train tasks of the path covers measured in time and kilometers. The composition changes occur when two train tasks are not on the same train sequence. The number of composition changes is kept low to ensure robustness of the path. The maintenance limitations and the limits on composition changes will be resources of the paths.

The paths of the different train units are generated iteratively. First a path is generated and then the network is updated. In practice this means that for each path found for a train unit the train tasks will already be covered by the train unit. If the train task has a fully covered node and all arcs connected to it will be removed from the network. Recall that at S-tog there are at most three train units in a composition. An overview of how to update the network with respect to covered demand in each node is as follows:

1. If the demand of the node is still greater than 0 there will be no cost associated with covering the node.
2. If the demand of the node is 0 (or less, in which case it will be rounded up to zero) there will be a cost associated with covering the node

3. If the composition of the node is of maximum length, the node is removed from the network i.e. it will not be available for future paths of other train units

7.1.4 Handling compositions when constructing shortest paths

During the shortest path generation phase we will construct paths feasible with respect to maintenance. Also, for all generated paths, we will attempt to meet requirements of feasibility with respect to composition changes. That is, the de-/coupling of train units from the composition depends on the order of the train units after each train departure from a depot. A train unit can only be decoupled at a depot if it is in the open end of a composition.

After a path for a train unit has been found, the network is updated and arcs are removed with respect to connection that are no longer feasible. Feasibility depends on the capacity of the individual train tasks.

Recall, that a composition change is feasible if we only remove or add train units in the open end of the composition and otherwise infeasible.

An example of infeasibility could be the following: We consider a train task, t_1 . The composition of t_1 consists of two train units, v_1 and v_2 , where v_1 is in the southern end of the composition and v_2 is in the northern end of the composition. Assume that the next task of v_1 after t_1 is the train task, t_2 . If t_1 arrives at a depot open for decoupling in the northern end, v_1 will be in the closed end and it can therefore not be decoupled. Hence, if this is the case the assignment of both t_1 and t_2 to v_1 is infeasible.

For each possible connection between two nodes in the network, $i \in V, j \in V$ the feasibility with respect to composition changes is ensured by a feasibility check. Whether a connection between i and j is feasible depends on the paths already passing the two nodes. In addition, it depends on the paths that we are currently constructing.

The information used for deciding whether a connection is feasible is:

Open end of composition : It is always possible to establish which end of a composition is open for decoupling. Generally, on a north depot the tracks are closed in the north end and open in the south end. The opposite applies to a south depot. All depots can be classified as either a north or a south

depot, even when they are intermediate on a train route. No depots allow de-/coupling in both ends of the composition.

Composition changes identified on train task knowledge : We always know if the addition of a node to a path imposes a composition change, namely because we know that a composition change only occurs when the train task considered as extension to a path is not on the same train sequence as the present last train task in the path. Because of this information we can unambiguously decide whether a connection is feasible with respect to composition changes.

If a composition is full with respect to maximum train length : Establishing whether a coupling is feasible with respect to composition is entirely a matter of knowing whether the composition will become too long.

If the open end is occupied : We update the network according to the chosen path after each run of the RCSPP. Therefore, we know for any node j considered as an extension to a path whether any of the train units assigned to j in earlier runs of the RCSPP will decouple after j and therefore must be in the open end.

The described process is carried out for all train sequences, see Algorithm 2 in Appendix A. The algorithm is used on all train units not assigned to a path in the first step of the construction heuristic. The complete process for feasibility check of compositions has not yet been constructed.

7.2 Improvement heuristic

The quality is heavily dependent on the order in which we consider the train units. We therefore expect that it will be necessary to use an improvement heuristic.

Given the solution found with the construction heuristic, the aim of the improvement heuristic is to gradually improve the solution by searching a solution neighborhood. We define our improvement heuristic by its neighborhood and the type of search heuristic chosen. Following are ideas for neighborhoods:

Swap of the tail end of two paths

We swap a subset of paths among train units. For example, if two train units, v_1 and v_2 , each have a path of tasks assigned, ρ_1 and ρ_2 . Let us assume that the

paths of the train units intersect at station $stat_1$, i.e. ρ_1 and ρ_2 both have some train task with arrival station $stat_1$. We divide ρ_1 and ρ_2 into two sub-paths each so that $\rho_1 \equiv \rho_{1a} \cup \rho_{1b}$ and $\rho_2 \equiv \rho_{2a} \cup \rho_{2b}$ where ρ_{1a} last train task has $stat_1$ as arrival station and ρ_{1b} has $stat_1$ as departure station. The same goes for ρ_{2a} and ρ_{2b} . Given that a swap feasible with respect to compositions can be made we swap the end paths of the train units so that v_1 is assigned to the path $\rho_{1a} \cup \rho_{2b}$ and v_2 is assigned to the path $\rho_{2a} \cup \rho_{1b}$.

Delete expensive path of a specific train unit

We choose a path, ρ , of a train unit, v , which is expensive, that is it either covers tasks of several train sequences or it covers one or more train tasks whose demand is covered with more train units than is needed relative to the expected passenger demands of the train tasks.

We remove v from ρ and update the network accordingly. Notice that when we remove a train unit from the compositions of a path, we will not violate any composition change constraints.

Finally, we reassign vacant train units to the train tasks of ρ which are insufficiently covered. In the process of finding a train unit, v' , we evaluate all train units, $v' \in \Upsilon$, disregarding the location of v' . If v' is not located at the correct departure depot relative to the departure depot of t , we try to find a route along train tasks earlier than t in the attempt to route the train unit to the departure depot of t before the departure time of t . Note that rerouting should maintain feasibility with respect to composition changes.

Choose a train task where the expected number of standing passengers is high

We choose a train task, t , for which the composition assigned to it offers an insufficient number of seats.

We then search all train units, $v \in \Upsilon$, to find one that can feasibly cover t . Again, as in the previously described neighborhood, we only consider train routes feasible with respect to composition maximum length and composition changes.

Choose a train task where its composition has many excess seats

We choose a train task, t , for which the expected demand is much lower than the number of offered seats in the composition.

One train unit is removed from the composition of t so that the number of seats

in the reduced composition fits the expected demand better. The removal of the unit must be feasible with respect to composition changes. For example, if there is an expected passenger demand on t of 298 passengers and the composition consists of a SA and a SE (resulting in a total of 486 seats), we try to remove the SE. If the removal of the SE is feasible with respect to composition changes, we remove the SA and try to reroute a SE train unit after the same process described for the two previous neighborhoods i.e. feasibly with respect to maximum composition lengths and composition changes.

7.2.0.1 Search heuristic

There are several possibilities for the choice of search heuristic e.g. Steepest Ascent Local Search, Simulated Annealing or Tabu Search are examples of procedures that can be chosen. A survey of search techniques for large-scale neighborhoods is available in Ahuja et al. [2002]. The choice of search heuristic still remain an open issue at the current stage.

7.3 Conclusion

As the RSRP is a recovery problem it is very important to find a feasible solution of a sufficiently good quality within sufficiently short time.

We expect a short running time of the first step of the construction heuristic, since it has a worst case complexity of $\mathcal{O}(|\Psi| \times |\Upsilon|)$ where Ψ is the set of train sequences and Υ is the set of train units. The running time of the second step of the construction heuristic is harder to predict. The running time depends on how many nodes are covered in the first step. Boland et al. [2006] give the worst case complexity of their generalized label setting algorithm of $\mathcal{O}(|A| \prod_{r+1}^R (\mathcal{W}_r + 1) 2^{|S|})$. Because of the first step of the construction heuristic we expect that we will get relatively good initial solution.

We have begun the implementation of the solution approach described in this chapter. We have finished the implementation of the first step of the construction heuristic and we are in the process of implementing the RCSP algorithm for the second step in the construction heuristic. We still lack the part of the implementation that concerns the feasibility of the composition changes. Additionally, the improvement heuristic also needs to be implemented.

Future research projects

In this chapter we discuss ideas for future research projects. We present ideas for projects which are extensions of the research regarding the RSRP and discuss projects which are complementary to the RSRP.

8.1 Column generation approach to the RSRP

The basic idea of column generation is to split the *Routing* model of the RSRP into a master and a subproblem. The specific problem characteristics are treated in the subproblem where feasible lines of work consisting of train tasks are generated according to train units. The master problem is a set partitioning model assigning train units to positions in train task compositions. Given information from the dual multipliers of the solution of the master problem columns are generated iteratively using the sub problem. An introduction to column generation is given in Desrosiers and Lübbecke [2005], Section 2.

The subproblem is a *Resource Constrained Shortest Path Problem*, abbreviated RCSPP. It can to a large extent be formulated as in Chapter 7. Feasibility with respect to composition changes is ensured by using as input a solution for the

underlying Position model¹. That is, the Position solution is used to construct the network for the RCSPP. It is from the *Position* model predefined how many train units and of which type must cover each train task. This is equivalent to the predefinition of how many nodes represent each train task. The arcs included in the network are those ensuring feasible composition changes.

The set partitioning formulation is used for other railway related problems. In Rezanova and Ryan [2006] the Train Driver Recovery problem (TDRP) is addressed. We will in the following argue that the advantageous qualities of the constraint matrix seen for the TDRP are also present in the *Routing* problem, given that the *Position* problem has been solved.

driver ₁				driver ₂				driver ₃				driver ₄			
1	1	1	1												
				1	1	1	1	1							
								1	1	1	1	1			
												1	1	1	1
1		1		1	1			1		1		1	1		
1				1	1		1	1		1	1	1			
	1	1				1	1	1	1			1		1	

Figure 8.1: Constraint matrix structure for the TDRP

In Fig. 8.1 the constraint matrix of the Master problem for the TDRP is illustrated.

The first horizontal block of rows, $*^1$, relates to the constraints for drivers. Each column relates to a driver. There can be only one 1 per row indicating that each driver can cover only one duty.

The second horizontal block of rows, $*^2$, relates to the tasks. Each column in the matrix refers to a specific train driver. If there is a 1 in a row, ρ , in the second block, horizontally, it means that the task corresponding to ρ is included in the duty indicated by the column where the 1 is located.

Besides the constraints referring to $*^1$ and $*^2$ the problem has a set of integrality constraints.

¹Recall that the Position model is described in Groth et al. [2008]

As argued earlier there is a strong resemblance between the structure of the RSRP and the structure of the TDRP. In Rezanova and Ryan [2006] it is further argued that the TDRP has the same structure as the crew rostering problem, reference is given to E. R. Butchers and Wallace [2001] and M. Gamache and Desrosiers [1999]. What is interesting about the latter resemblance of problems is that the crew rostering problem possesses strong integer properties because of the underlying perfect matrix structure of the submatrix corresponding to each driver, see Ryan and Falkner [1988]. Rezanova and Ryan [2006] proves that the TDRP also possesses strong integer properties. The strong integer properties will mean a potentially shorter solution time. Therefore, we are interested in whether this also applies for the RSRP.

The immediate structure of the RSRP is the same as that of the TDRP. They are set partitioning models where duties are generated in a RCSPP. This means that the duties generated in the RCSPP and represented by a parameter in the set partitioning model are feasible with respect to a large set of constraints such as e.g. maximum length in time or arrival station of last task in the duty.

The difference between the TDRP and the RSRP is that in the RSRP each line referring to the \ast^2 constraints in 8.1 refers to a position in a train task. This means that each train task in the timetable can be represented by more than one row depending on the composition found for the *Position* problem. However, as no constraints relate to the positions in train tasks in the set partitioning formulation they are simply to be seen as an extension of the set of tasks that must be covered. Hence, each position in a composition is the same as a task. As there are no interdependencies between positions within the composition of the same train task, the structure as known from the TDRP is maintained.

8.2 Rolling stock depot recovery

Closely related to the RSRP is the recovery of the rolling stock depot plans. At the rolling stock depots rolling stock is driven to and from depot tracks during the operation. As the depot capacity is limited on the S-tog rolling stock depots, it is necessary to adapt the depot plans to a rolling stock schedule. It is possible that no feasible depot plan can be found for a given rolling stock plan. Is this the case, the rolling stock schedule must be altered.

The depot planning problem can be partitioned into two main phases. First, a matching of arrivals and departures must be formed. A train unit arriving at the depot must be matched with some later departure from the same depot. Second, train units must be parked at specific depot tracks in such a way that

when each train unit is to leave the depot again, it is in the open end of the depot track. The input to a depot planning process is the dimensions of the depot and specifications of train units such as e.g. their length.

A real time disruption in the rolling stock schedule will affect the depots on the routes of the affected train units if there are couplings or decouplings on the disrupted departures. One effect can be that a train unit, v_c , is not coupled to its departing train at the scheduled time and is therefore blocking for other train units at the same depot track which are departing from the depot at a later point in time than v_c . Another effect can be that the train unit, v_d , is not decoupled at the scheduled time. If v_d or some other train unit is decoupled at a later point in time a new feasible parking must be found which does not block other departures from the depot. Notice that for a subset of train units the original parking location may still be feasible.

In the recovery process each depot with a disrupted depot plan is handled separately. A possible depot recovery procedure is to base any recovery solution on the existing solution making a minimum of alterations. In practice a limited recovery period will be considered i.e. a recovery plan is derived by which the operation returns to normal within the recovery period. The recovery problem can at first be limited to those tracks which are directly affected by the disruption and those tracks which at no time is filled up during the recovery period. If no solution can be found, the recovery period can be increased or the number of depot tracks considered can be expanded.

For the depot recovery problem, feasibility is more important than optimality. One reason for this is that the different solutions will vary little in operational cost.

For an introduction and references on the tactic depot planning problem see Føns [2006].

8.3 Evaluation of robustness of a RSR solution by simulation

There are two main applications of a simulation model built for simulating the RSRP. The first application is the evaluation of the rolling stock schedule at different planning phases with respect to robustness. The second is off line training of the rolling stock dispatchers. In this section we will concentrate on the first application.

In general we consider coupling and decoupling of train units as the reason why rolling stock schedules become non-robust. With a simulation tool it can be investigated whether the couplings and decouplings planning makes the rolling stock schedules less robust.

A rolling stock scheduling simulation model can be more or less detailed. Immediately we identify two level of details. The first level only includes the depots as stations where trains arrive and depart and units are coupled to or decoupled from the train without including the details of the parking at the rolling stock depot. Coupling and decoupling are hence included with an estimated duration time and a distribution of delays that can be anticipated on the execution of the coupling and decoupling activities. The second level of detail is an expansion of the first level where the details of the individual depots are included and the traffic of train units within the depot is included as well.

The simulation model can either be formulated microscopically or macroscopically. Microscopically the timetable, de-/acceleration and weight of train unit types driving and all railway network details are included i.e. tracks, track connections, platforms, switches, signalling system etc. Macroscopically we consider the effects of the railway network from a time schedule perspective i.e. we know the timetable, the headways between trains, the minimum driving time between stations, the time it takes for a depot driver to walk from the parked train back to the platform and similar. When there is no necessity to be precise but only a relative evaluation of a rolling stock schedule's ability to maintain punctuality is wanted, a macroscopic simulation model will likely be sufficient. The first level of detail described above demand a macroscopic model. The second level can be handled with either a micro- or a macroscopic model.

It is possible to construct a macroscopic simulation model which has the same level of detail as the simulation model presented in Hofman et al. [2006]. Stations and rolling stock depots can be formulated generically including the complete network in one model. However, if the objective is to investigate the shunting process it is preferable that each rolling stock depot is modelled and evaluated separately.

Conclusion

In this thesis we have discussed the area of railway planning and in particular rolling stock dispatching within the area of railway dispatching. All work is based on the railway network and operation of S-tog. We have concentrated on the area of railway passenger transport and within rolling stock planning the area of self-propelling multiple train units.

Automatic decision support is a key contribution to the competitiveness of a railway operator in all planning phases. The use of decision support and advanced, automatic planning systems is diffusing among the railway operators of Europe. One of the planning phases now receiving increased attention is the real-time planning phase. Real-time planning demands valid data and stable system integration. Furthermore, methods used for solving the real-time problems must supply solutions of adequate quality within a sufficiently low computation time.

9.1 Main contributions

The main theoretical contributions are in the four papers enclosed in the thesis.

In collaboration with researchers at NSR we have written a survey of the main problems and their interdependencies within disruption management in pas-

senger railway transportation. The survey lists the few works that have been done so far within the subject and furthermore gives a clear indication of the challenges still remaining.

The ability to make immediate recovery lies also in the dedicated planning of timetable and schedules of resources with respect to robustness. We have completed a robustness study where a macroscopic simulation model was developed. Results show that there is an upper limit on the size of buffer times in timetables that will be beneficial to the punctuality. Also, small delays have large knock-on-effect when one larger delay has occurred. Finally, the study gave a clear view on which recovery strategy to employ for different sizes of delays.

The results achieved by the simulation model have initiated a follow up project assessing the robustness of two new timetables of which one was chosen to operate.

We have developed a mathematical model for reinsertion of a cancelled train line. The *Reinsertion* model is a Mixed Integer Programming model handling a practical problem indirectly with only few details on the specific train routes. Regardless, it is fully applicable in the operation. It forms the basis for a decision support tool which is today used in the operation at S-tog. The *Reinsertion* model is formulated on the basis of the reinsertion problem of S-tog. However, it is applicable for other railway operators who have a fully periodical timetable and relatively short circuits of train lines.

The impact of the Ph.D. project can already be seen in S-tog. The *Reinsertion* model is used in the operation. It is the first real-time optimization tool and has helped to bring the recovery process into focus. Also, it has improved the work process for the rolling stock dispatcher. Finally, it has lead to a higher reliability which directly affect the service level and improves the customer satisfaction level.

A decomposed model for the RSRP has been constructed. The decomposed approach is partitioned in the *Position* model, the *Sequence* model and the *Routing* model. The *Position* model assign train unit types to train tasks under consideration of the train units position in the composition of each train task. The *Sequence* model assign train units to train sequences. The *Routing* model assigns train tasks to specific train units.

The decomposed model is one of the first contributions aiming at supporting rolling stock recovery and it considers the complete recovery process from finding the compositions for each train tasks to routing of the physical train units. The model have greatly enhanced our knowledge on the RSRP. We now know that the *Routing* model is extremely sensitive to the solution found in the *Position* model

and that the *Sequence* model is of significant importance to the applicability of the decomposed approach. Most importantly, the decomposed approach shows promising computational results regarding the general quality of solutions. Also, the observed computational times are sufficiently low for real-time applicability.

9.2 Further developments

There are practical and organizational projects that must be implemented before a decision support tool for the rolling stock recovery problem is functional.

Practically there is a project of preparing data and ensuring the right system integration. These are major IT projects and the success of a rolling stock recovery decision support tool depends on their successful completion. In an ongoing project in S-tog data from each train is being accessed in real-time ensuring accurate data on where train units are located. In parallel with the implementation of a decision support tool sufficient integration with existing system must be ensured e.g. a rolling stock recovery tool must be able to pass on information on the changes it suggest to a rolling stock surveillance system.

When designing and introducing new system tools there is an immediate challenge in adjusting the known processes of the employees according to the improved process imposed by the tool (and vice versa). This adjustment of the processes can be a part of a change management project meant to ensure the successful delivery of the finished system to the end user.

9.3 Summing up

The thesis shows potentials for the use of decision support tools in rolling stock dispatching. A decision support tool has been developed and implemented in practice and a model for another decision support tool show promising results regarding applicability.

The thesis has also produced ideas for related research projects as those described in Chapter 8. Specifically must be emphasized that a system for rolling stock depot recovery is crucial to the applicability of a rolling stock recovery system.

The scientific problem of rolling stock recovery is one in a series of real-time decision support problems which needs further research before advanced auto-

matic decision support tools can be implemented and used in the operation. The immediate challenge today is to develop scientific formulations and methods which can be candidate for the core of future decision support systems. These methods must be viable with respect to quality and computation time.

APPENDIX A

Generalized Label Setting Algorithm

We here describe the Generalized Label Setting Algorithm as presented in Boland et al. [2006]. We have added one modification in line 15 of Algorithm 2, namely, the feasibility check of connections.

Name	Description
V	The set of nodes
A	The set of arcs
$G = (V, A)$	The graph
$i \in V$	A node in the graph
$s \in V$	The source node of the graph
$t \in V$	The sink node of the graph
L_i	The set of labels on node i
T_i	The set of treated labels on node i
P_t^k	The path on the sink corresponding to label k
\mathcal{P}_*	The final set of Paths on the sink
W_i^k	Vector of resource usages of path denoted by label k in node i
C_i^k	Cost of path denoted by label k in node i

Algorithm 2 Generalized Label Setting Algorithm

```

1: Step 0: Initialization
2: Set  $L_s = (0, 0)$  and  $L_i = \emptyset$  for all  $i \in V \setminus \{s\}$ .
3: Initialize  $I_i$  accordingly and set  $T_i = \emptyset$  for each  $i \in V$ .
4:
5: Step 1: Selection of the label to be treated
6: if  $\bigcup_{i \in V} (I_i \setminus T_i) = \emptyset$  then
7:   stop; all efficient labels have been generated,
8:   Return  $\mathcal{P}_* = \{P_t^k : k \in I_t\}$ , the set of paths corresponding to labels on  $t$ .
9: else
10:  Choose  $i \in V$  and  $k \in I_i \setminus T_i$  so that  $W_i^k$  is lexicographical minimal.
11: end if
12:
13: Step 2: Treatment of label  $(W_i^k, C_i^k)$ 
14: for all  $(i, j) \in A$  do
15:   if  $\text{CreateNewLabel}(i, (W_i^k, C_i^k), j, (W', C'))$  and  $\text{Feasible}(i, j)$  then
16:     if  $(W', C')$  not dominated by any label in  $L_j$  then
17:       Set  $L_j = L_j \cup \{(W', C')\}$ ,
18:       Remove any dominated labels from  $L_j$ ,
19:       Update  $I_j$ 
20:     end if
21:   end if
22: end for
23: Set  $T_i = T_i \cup \{k\}$ 
24: goto Step 1

```

```

1:  $\text{CreateNewLabel}(i, (W, C), j, (W', C'))$ 
2: if  $W^r + \overset{r}{w}_{ij} + \overset{r}{w}_{jt} > \mathcal{W}_r$  for some  $r \in 1, \dots, R$  then
3:   Return FALSE
4: else
5:   Set  $W' = W + w_{ij}$  and  $C' = C + c_{ij}$ 
6:   Return TRUE
7: end if

```

APPENDIX B

Disruption Management in Passenger Railway Transportation

IMM-Technical Report-2007-03.

Disruption Management in Passenger Railway Transportation

Julie Jespersen Groth^{a,b}, Daniel Potthoff^{c,e}, Jens Clausen^{a,b},
Dennis Huisman^{c,e,f}, Leo Kroon^{d,e,f}, Gábor Maróti^{d,e}, and Morten
Nyhave Nielsen^a

^a DSB S-tog, Kalvebod Brygge 32, 5., 1560 Copenhagen V., Denmark

^b Department of Informatics and Mathematical Modelling Technical University of
Denmark, DK-2800 Kongens Lyngby, Denmark

^c Econometric Institute, Erasmus University Rotterdam P.O. Box 1738, NL-3000 DR
Rotterdam, The Netherlands

^d Rotterdam School of Management, Erasmus University Rotterdam P.O. Box 1738,
NL-3000 DR Rotterdam, The Netherlands

^e Erasmus Center for Optimization in Public Transport (ECOPT)

^f Department of Logistics, Netherlands Railways P.O. Box 2025, NL-3500 HA
Utrecht, The Netherlands

Abstract

This paper deals with disruption management in passenger railway transportation. In the disruption management process, many actors belonging to different organizations play a role. In this paper we therefore describe the process itself and the roles of the different actors.

Furthermore, we discuss the three main subproblems in railway disruption management: timetable adjustment, and rolling stock and crew re-scheduling. Next to a general description of these problems, we give an overview of the existing literature and we present some details of the specific situations at DSB S-tog and NS. These are the railway operators in the suburban area of Copenhagen, Denmark, and on the main railway lines in the Netherlands, respectively.

Since not much research has been carried out yet on Operations Research models for disruption management in the railway context, models and techniques that have been developed for related problems in the airline world are discussed as well.

Finally, we address the integration of the re-scheduling processes of the timetable, and the resources rolling stock and crew.

B.1 Introduction

Many Europeans travel frequently by train, either to commute or in their leisure time. Therefore, the operational performance of railway systems is often discussed in the public debate. Travelers expect to arrive at a specific time at their destination. If they travel by rail, they expect to arrive more or less at the time published in the timetable. However, unforeseen events often take place, which cause delays or even cancelations of trains. As a result, passengers arrive later than expected at their final destinations. Due to missed connections, the delay of a passenger can be even much larger than the delays of his individual trains.

Due to the importance for the public on one hand and the deregulation of the railway market on the other, railway operators now put more emphasis on their operational performance than in the past. Furthermore, due to the separation of the management of the infrastructure and the operations in many European countries (including Denmark and the Netherlands), several organizations are responsible for the performance of the railway system.

This paper deals with passenger railway transport only. However, in addition to the passenger railway operator itself, the infrastructure manager and other (also cargo) operators have a strong influence on the performance of the railway services of that single operator. Therefore, the role and the objectives of the infrastructure manager and of the operators are discussed.

We consider two passenger railway operators in more detail: DSB S-tog and NS. DSB S-tog is the operator of local train services in the greater Copenhagen area, see Figure B.1. NS is the main operator in the Netherlands, having the exclusive right to operate passenger trains on the so-called Dutch Main Railway Network until 2015, see Figure B.2. Both companies operate a set of lines on their network, where a *line* is defined as a route between two stations operated with a certain frequency, e.g. line A of S-tog runs between Hillerød and Hundige every 20 minutes.

Unfortunately, trains do not always run on time due to unexpected events. Examples are infrastructure malfunctions, rolling stock break downs, accidents, and weather conditions. Such events are called *disruptions*. To give an indication, the numbers of disruptions related to infrastructure in the Netherlands during the first half of 2006 are reported in Table B.1.

Table B.1 shows that the Dutch railway network has approximately 22 disruptions related to the infrastructure per day with an average duration of 1.7 hours. Note that disruptions caused by the operators, e.g. rolling stock break downs and crew no-shows are not reported in this table. The proportion between the

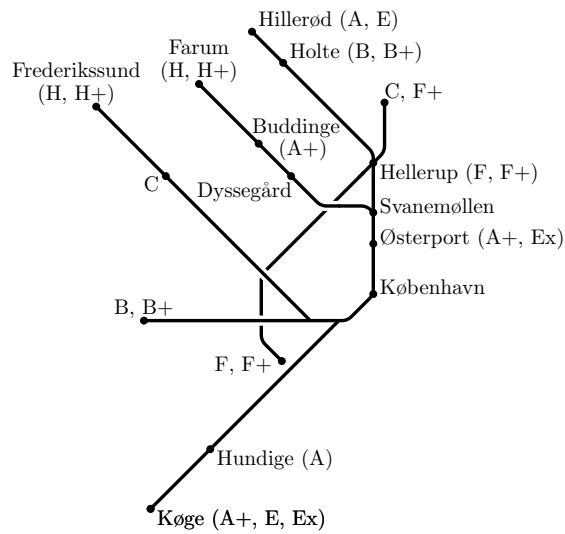


Figure B.1: The S-tog railway network

disruptions caused by the operators and the infrastructure is roughly 50-50 in the Netherlands.

Different information is recorded for S-tog. Table B.2 shows the number of affected trains in an average month for 2006. An affected train is either at least 2.5 minutes late on departure or canceled. Table B.3 further details the information regarding that part of the affected trains where the disruption is contributed to S-tog.

Class	Disruptions	Avg. duration	Total duration
Technical failure	1656	2.2	3680
Third parties	1471	1.0	1491
Weather	172	2.3	393
Others	693	1.7	1208
Total	3992	1.7	6772

Table B.1: Disruptions in the Netherlands related to infrastructure during the first half of 2006 (ProRail [2006])

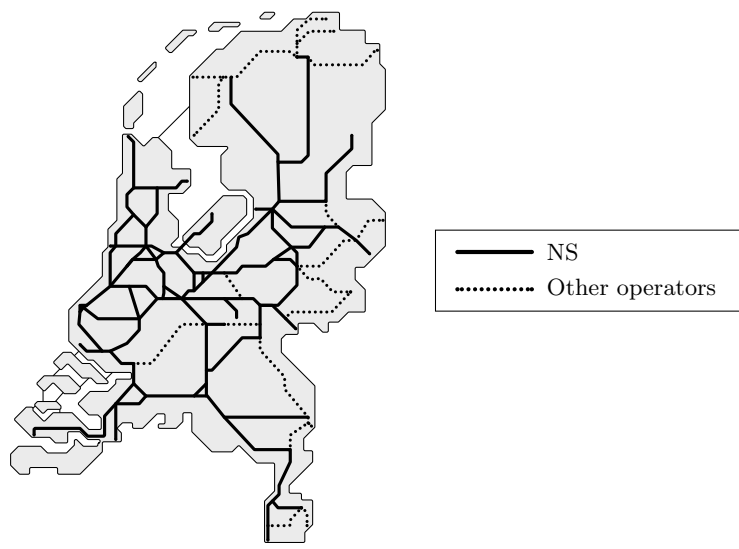


Figure B.2: The Dutch railway network (in 2005)

Of course, infrastructure managers and operators try to avoid disruptions. Unfortunately, many of them are hard to influence. Therefore, it is very important to limit the consequences of these disruptions. A very common problem in railways is that, due to the strong interdependencies in the railway network and due to cost efficient resource schedules, disruptions are very likely to spread over the network in space and time. This well-known phenomenon is called *knock-*

Responsible	Infrastructure manager	S-tog	Externally caused
Affected trains	4746	3981	660

Table B.2: Disruptions in the S-tog traffic for an average month in 2006 subdivided according to responsability.

Responsible	Rol. St.	Drivers	Dispatch.	Maint.	Pass.	Misc.
Affected trains	1131	665	88	44	1737	316

Table B.3: Disruptions contributed to S-tog for an average month in 2006 (in total 3981) subdivided according to cause.

on effect. The key to a good performance of railways is to limit the knock-on effect and thereby to limit the impact of single disruptions. Therefore, effective disruption management is required.

So far, Operations Research (OR) models have hardly been applied in practice for disruption management in railway systems. Nevertheless, it is our strong belief that OR models can play an important role to limit the impact of disruptions and thereby to improve the performance of railway systems. This belief is supported by the fact that nowadays OR models and techniques play a major role in several railway companies during the *planning* phase, where the focus is on a good balance of the service level offered to the passengers and efficiency of the resources rolling stock and crew. For an overview on these models and techniques, we refer to surveys of Assad [1980], Cordeau et al. [1998], and Huisman et al. [2005]. Moreover OR models have proven to be quite effective already for supporting disruption management processes in the airline context, see e.g. Yu et al. [2003].

The objectives of this paper are twofold. First, we intend to give a comprehensive description of the problems arising in disruption management for railway systems. Second, we aim at attracting new researchers to this field by describing the challenges that railway companies are faced with to improve their operational performance.

The remainder of this paper is organized as follows. In Section B.2 we give a description of disruption management for railway systems, including a description of organizations and actors involved in this process. In Sections B.3-B.5, we discuss timetabling, rolling stock and crew aspects of the disruption management process. Section B.6 deals with the advantages and possibilities of integrating some of these processes. Finally, we finish the paper with some concluding remarks in Section B.7.

B.2 Description of disruption management

Clausen et al. [2005] give the following definition of a *disruption* in relation to airline operations: "An event or a series of events that renders the planned schedules for aircraft, crew, etc. infeasible." By definition, a disruption is hence a cause rather than a consequence. In this paper we use the same definition for railway operations, substituting "aircraft" with "rolling stock".

A disruption does not necessarily have immediate influence on the timetable - some disruptions like a track blockage renders the planned timetable immedi-

ately infeasible, while others as e.g. shortage of crew due to sickness may lead to cancelations either immediately, in the long run or not at all, depending on the amount of stand-by crew. Note that a disruption leads to a *disrupted situation*. Even though this is a slight abuse of terms, we will occasionally refer to the disrupted situation as the disruption itself.

Accordingly, we define *railway disruption management* as the joint approach of the involved organizations to deal with the impact of disruptions in order to ensure the best possible service for the passengers. This is done by modifying the timetable, and the rolling stock and crew schedules during and after the disruption. The involved organizations are the infrastructure manager and the operators.

Of course, one first has to answer the question if the situation is disrupted, i.e. if the deviation from the original plan is sufficiently large or not. Similar to the airline world (see Kohl et al. [2004]), this question is normally answered by dispatchers monitoring the operations. In the railway world, however, it seems to be more difficult to judge an overall situation, even for experienced dispatchers. The latter is in particular true in case of a dense railway system. In the reminder of this paper, this issue is not considered further.

In Section B.2.1 we define terms enabling us to describe and discuss capacity issues in railway networks. The Sections B.2.2 to B.2.4 introduces a framework of organizations, actors and processes in disruption management, which is valid for several European railway systems. In Section B.2.5 we discuss the organizational context of the disruption management process and in Section B.2.6, we describe a number of issues that are related to disruption management, such as robust planning.

B.2.1 The capacity of a railway network

The state of the daily operation of a train operator at some point in time is influenced by a number of factors, including the current state of the infrastructure (the rail network), and the state of all resources necessary in the operational phase, most notably rolling stock and crew. In the following we introduce the concepts of infrastructural capacity, operational capacity, utilization, and residual operational capacity.

The *infrastructural capacity* $IC(t)$ of a rail network N in a particular state is the maximum amount of traffic which is continuously able to flow through N in this state. The state may be described by the status of a number of parameters

as e.g. the set of available tracks and for each track the state with respect to signals, and the maximum allowed speed for each track segment. Note that $IC(t)$ is independent of the current amount of traffic. The maximum value of $IC(t)$ over all possible states is sometimes referred to as the capacity of N .

At any point t in time, the network N and the resources are in one of their possible states. The *operational capacity* $OC(t)$ of the network is the maximum amount of traffic which is continuously able to flow through N with the current states for network and resources. Note that $OC(t)$ is always less than or equal to $IC(t)$ for N in the current network state - one can never run more traffic than the infrastructural capacity allows for, but may not have resources enough to utilize this completely.

The *utilization* $U(t)$ of the network at time t is the amount of continuously flowing traffic in the network N at time t . Through the operational capacity of the network, $U(t)$ is depending on both the network state and the state of each resource. Note that a number of feasible values for utilization exist for each set of states for the network and resources. $U(t)$ is always less than or equal to $OC(t)$. The *residual operational capacity* or just the *residual capacity* at time t is now the difference between the operational network capacity and the utilization at time t : $R(t) = OC(t) - U(t)$.

The states of the network and each of the resources are dynamic. The states are influenced by planned actions as inserting or taking out rush-hour trains, new crew meeting in, and trains taken out for maintenance. However, the states are also influenced by disruptions as e.g. engine break downs, inserting stand-by crew or rolling stock, or taking out train lines. A disruption typically decreases either the operational capacity, the utilization or both, while a recovery action typically increases either the residual capacity, the network utilization, or both. Since the utilization is less than or equal to the operational capacity, a decrease in operational capacity can never lead to an increased residual capacity.

Increasing the residual capacity may be achieved e.g. by decreasing utilization (e.g. canceling trains or entire train lines). Note that this operation does not necessarily increase the operational capacity. The state of the system may be changed to a state with larger operational capacity by e.g. allowing trains to run faster, decreasing the turn-around times at end stations, or inserting stand-by resources. This does not automatically increase the utilization.

Finally note that a recovery action in general serves two distinct but often conflicting purposes: Increasing the network utilization, and changing the states of the resources to more preferable states. Canceling a train is very good from the resource perspective in that the action increases the residual capacity of the network as well as the available amount of both crew and rolling stock,

and thereby possibly the operational capacity, but at the cost of a decreased utilization. Moreover, the “goodness” of a particular state may be difficult to quantify. For example, a state at time t is usually considered good, if it is “close” to the planned state.

B.2.2 Organizations

The organizations directly involved in disruption management are the infrastructure manager and the railway operators. These organizations usually have contracts with the involved government. Moreover, they have a certain relationship with each other. These issues are described below.

The infrastructure manager has a contract with the government that obliges it to provide the railway operators with a railway network of a certain infrastructure capacity and reliability. The infrastructure manager has also the responsibility of maintaining the railway network as efficiently as possible.

A passenger railway operator obtains from the government a license to operate passenger trains on the network. The operator is contractually bound to provide a performance that exceeds certain specified thresholds on certain key performance indicators. For example, there may be thresholds for the number of train departures per station, for the (arrival) punctuality at certain stations, for the percentage of caught connections, for the seating probability, etc. Here, the punctuality is the percentage of trains arriving within for example 3 or 5 minutes of their scheduled arrival time at certain stations. The realization figures on these performance indicators have to be reported to the government periodically. If an operator does not reach one of the thresholds, it has to pay a certain penalty to the government. If the performance is very poor, another operator may be given the license to operate trains on the network.

As a consequence, usually the main objective of the railway operator is to meet all thresholds set in the contract with the government at minimum cost. The latter is due to the fact that the railway operators are commercially operating companies. Thus the number of rolling stock units on each train must match with the expected number of passengers. Deadheading of rolling stock units between depots and to and from maintenance facilities must be minimized. Furthermore, the number of crews needed to run the operations and to cover unforeseen demand must be minimized as well.

In more detail, an important objective of the operators in the disruption management process is to minimize the number of passengers affected by the disruption,

and to minimize the inconvenience for the affected passengers. Indeed, small delays of trains are usually not considered as a bad service by the passengers, but large disruptions are. If passengers are too often confronted with large disruptions, which usually lead to long extensions of travel times and, even worse, to a lot of *uncertainty* about travel options and travel times, they may decide to switch to a different mode of transport. In relation to this, passenger operators usually prefer to return to the original timetable as soon as possible after a disruption. Indeed, the original timetable is recognizable for the passengers. Therefore, the original timetable provides a better service than a temporary ad hoc timetable during a disruption.

The passengers are the direct customers of the railway operators, and they are only indirect customers of the infrastructure manager. This may imply that the manager has less knowledge of the expected passenger demand on each train and of the real-time passenger locations in the operations. The latter may prohibit a passenger focused dispatching, and may instead lead to a network capacity focused dispatching, i.e. dispatching focusing on supplying sufficient buffer times in the network to recover from disruptions.

Furthermore, each delay of a train may be attributed either to a railway operator or to the infrastructure manager, depending on the nature of the disruption. However, this creates a natural conflict between the organizations that may prohibit an effective communication and co-operation in the operations. The latter may be counter-productive for the operational performance of the railway system. Thus, although the infrastructure manager and the railway operators have the same general objective of providing railway services to the passengers of a high quality level, there are also conflicting elements in their objectives.

B.2.3 Actors

In railway disruption management, the actors are the dispatcher of the infrastructure manager and those of the railway operators. The major tasks to be carried out are *timetable adjustment*, *rolling stock re-scheduling*, and *crew re-scheduling*. Figure B.3 shows how the responsibilities for the different elements are shared among the actors.

The infrastructure manager controls and monitors all train movements in the railway network. *Network Traffic Control* (NTC) covers all tasks corresponding to the synchronization of the timetables of the different operators. NTC has to manage overtaking, re-routing, short turning, or canceling trains in order to prevent them from queueing up. The latter is a permanent threat at the basically

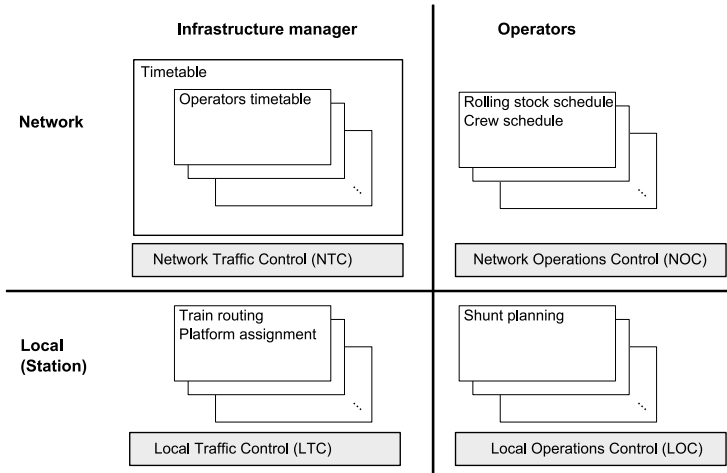


Figure B.3: Schematic view of actors, timetables and resource schedules

one-dimensional railway infrastructure. Queueing up of trains immediately leads to extensions of travel times.

On a local level, the process is managed by the *Local Traffic Control* (LTC). For example, LTC is responsible for routing trains through railway stations and for platform assignments. Safety is ensured by headways and automatic track occupancy detection systems.

The *Network Operations Control* (NOC) of each passenger operator keeps track of the operations of the operator on a network level. The dispatchers of NOC are acting as decision makers for the operator in the disruption management process. Depending on the size of the operator, there is one or more dispatchers for rolling stock and crew, respectively. These dispatchers monitor and modify the rolling stock and crew movements. NOC dispatchers are the counterparts of the dispatchers of NTC.

Dispatchers of the *Local Operations Control* (LOC) of the railway operators are responsible for coordinating several local activities at the stations, such as shunting processes. They support NOC by evaluating whether changes to the rolling stock schedules can be implemented locally.

Train drivers and *conductors* are also important elements in the disruption management process. They are usually the first ones that are confronted with passengers that are affected by a disruption. If train drivers and conductors work on different lines, they may carry a delay from one line to another. In order

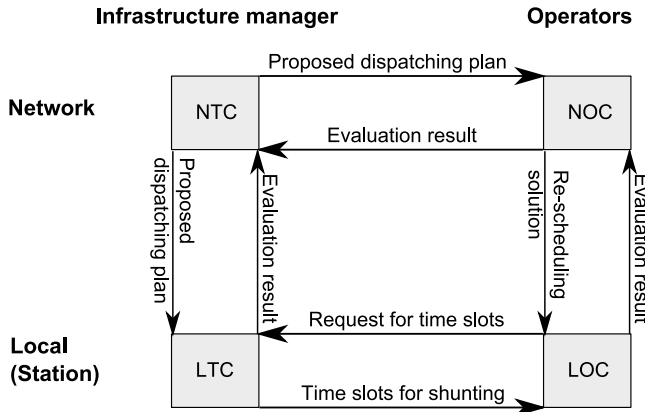


Figure B.4: Information flow during the dispatching plan development

to avoid this situation, the crew dispatchers may have to modify several duties. Besides making the decisions, the dispatchers also have to instruct and sometimes to convince the crew members to carry out the modifications, see Section B.5.

B.2.4 Processes

NTC dispatchers constantly monitor the operations and have to decide if an actual situation is a disruption or will lead to a disruption in the near future. When this is the case, they start the disruption management process. Within this process, the original timetable may need to be changed. This is done by carrying out a dispatching plan. Figure B.4 displays the information flows between the different actors in this process.

First, NTC determines all trains that are affected by the disruption. NOC of the corresponding operators must then be informed about the disruption and its direct consequences. In the next step, the dispatchers have to find out to which extent it is still possible to run traffic on the involved trajectory. Some pre-defined emergency scenarios give an indication about which trains should be overtaken, re-routed, short turned, or canceled. Using this information, an initial dispatching plan can be constructed. This dispatching plan must be evaluated by LTC. Almost simultaneously, the proposed dispatching plan is communicated to NOC of the operators. A complicating factor is the uncertainty about the duration of the disruption, for example NTC can only *estimate* how long it will take to repair a broken switch or signal.

The dispatching plan may correspond to changes in the planned operations of several operators. As a whole, these changes are compatible with respect to the safety regulations. However, for the operators it may be impossible to operate the dispatching plan due to their resource schedules for rolling stock or crew. Therefore, the decision about the dispatching plan is taken in consultation between the infrastructure manager and the operators.

Hence, NOC dispatchers have to check whether it is possible for them to operate the proposed dispatching plan. In particular, they have to check whether they can adapt their resource schedules to the proposed dispatching plan. Furthermore, LOC has to verify that the modified timetable and the adapted resource schedules can be carried out locally. Because of the combinatorial nature of the resource schedules and the limited time available, not all re-scheduling options can be evaluated. The re-scheduling solutions represent a trade-off between the available time and the quality of the solution.

This evaluation procedure can basically have three different outcomes. First, NOC and LOC may find a re-scheduling solution to the proposed dispatching plan where no additional cancelations or delays are needed. Second, they may find an initial solution, but trains have to be canceled in a second stage because rolling stock and/or crews are unavailable. A cancelation of a train has, however, a strong negative impact on the service level. Finally, NOC may come up with a request for changes to the proposed dispatching plan if this enables them to construct a much better solution.

Of course, not only one but several operators may ask for changes in the proposed dispatching plan. When these requests are conflicting, it is the responsibility of NTC to make a fair decision. This may involve another iteration of proposal and evaluation between NTC and the operators.

After the final decision about the dispatching plan has been taken by NTC, it is communicated to LTC and to the operators. LTC has to implement the new train routes and to change platform assignments. NOC has to inform the train drivers and conductors whose duties have been changed. LOC has to generate new shunting plans. LOC communicates directly with LTC to ask for time slots for shunting movements in the station area. Furthermore, passengers need to be informed in trains, at stations, and via internet and teletext about the changes in the timetable and alternative travel routes.

B.2.5 Organizational issues

The description in Section B.2.3 of the actors in the disruption management process is a functional description, and not an organizational. For example, it suggests that all dispatchers of each of the mentioned actors are located in the same office. However, this need not be the case.

For example, in the Danish case, NTC, LTC and the timetable and rolling stock dispatcher of the NOC of S-tog are located in the same room, but the crew dispatcher of NOC is located at the crew depot of S-tog. This division was made on request of the train drivers. In practice, it creates some challenges regarding effective communication between the different dispatchers.

In the Netherlands, the situation is even more complex: the Netherlands have been split up into 4 regions, and each region has its own NTC office and its own NOC office of NS. Moreover, there is a central NOC office of NS for coordinating the rolling stock re-scheduling process. Similarly, there are 13 LTC offices and 13 LOC offices of NS. Obviously, this organizational split leads to a lot of additional communication within NTC and within NOC, which is counter-productive in the disruption management process. Therefore, there are currently plans to bring all offices of NTC together, and to do the same with the NOC offices. Moreover, it is investigated how the separation between the infrastructure manager and the operators can be reduced.

B.2.6 Related fields

Delay management is closely related to disruption management. Consider the following situation, typical for railway systems. For a passenger, even a small delay of a train can increase his travel time by 20, 30 or 60 minutes if he misses a connection and has to wait for the next train. A similar situation exists for air traffic within a hub-and-spoke network when a flight arrives late at a hub. When the delay of a feeder train is not too large, it is possible to keep connections for passengers alive by delaying the departures of connecting trains a few minutes. The delay management problem is to find optimal wait-depart decisions for connecting trains such that the sum of the passenger delays is minimized. By keeping connections for passengers alive, an important criterion contributing to the service level of a railway system is addressed, namely the passenger satisfaction.

The wait-depart decisions correspond to minor changes to the original plans. The difference to disruption management is that, in delay management, it is

usually assumed that the changes to the timetable can be conducted without re-scheduling rolling stock and crew, see Schöbel [2004]. However, the decisions are taken by dispatchers of NTC and NOC that are also involved in disruption management, see Section B.2.3.

Another related issue is *robust planning*. Robust planning aims at making timetables and resource schedules less sensitive to disruptions. Robust planning approaches are called *pro-active*, since they take disruptions into account prior to their appearance.

There are two ways of interpreting robustness. The first one is to consider a plan robust if disruptions can be absorbed or the resulting knock-on effects can be reduced. We denote this property of a plan as the *absorbing capacity*. The second way of interpreting robustness is to consider a plan robust if it is well suited for recovery in case of disruptions. This property is called the *recovery capacity* of the plan.

The absorbing capacity of a plan is increased by introducing buffer times and by avoiding certain undesirable structures, such as short headways between trains, for which it is known that they are likely to propagate delays. Plans with a high absorbing capacity can compensate small disruptions completely, and they can reduce the consequences of larger ones. However, the high absorbing capacity usually comes at a price in terms of an increased cost of the planned operation.

Recovery capacity oriented robust planning is seeking plans that work well under one or several recovery strategies. Most recovery strategies use recovery actions that rely on certain desirable structures in the original plans. For example, initially planned crew connections can be swapped in the operations. In order to increase the recovery capacity of a plan, one tries to include such swapping options sufficiently and at the right locations in the plan. Moreover, plans are easier to recover when drivers and conductors stay together during their complete duty (the concept of *train teams*), and with the rolling stock. In a disrupted situation, adequate recovery strategies are easier to find when the recovery capacity of the plan is high.

Several methods have been proposed in order to increase the absorbing capacity of timetables. See Huisman and Boucherie [2001], de Kort [2000], Middelkoop and Bouwman [2000], Soto y Koelemeijer et al. [2000], and Kroon et al. [2005] for recent developments in this area.

In order to create rolling stock circulations that less likely propagate delays, railway operators use planning rules based on experience. For example, the rolling stock circulations of NS are planned on a line-by-line basis and, preferably, each line is operated by a single rolling stock type, see Huisman et al. [2005]

and Fioole et al. [2006]. No research has been done yet on more sophisticated methods for robust rolling stock planning.

Research on sophisticated methods for robust crew scheduling has so far only been done in the airline context. We refer to Ehrgott and Ryan [2002], Schaefer et al. [2005], and Yen and Birge [2006] for methods to increase the absorbing capacity, and to Shebalov and Klabjan [2006] for a method to increase the recovery capacity.

Stand-by rolling stock and crew planning are also interesting issues in the context of disruption management. During the planning phase, the number of stand-by rolling stock and crew and their positions have to be determined. To the best of our knowledge, this problem has not been addressed in the railway literature yet. A first reference dealing with a similar problem in the airline context is Sohoni et al. [2006].

B.3 Timetable adjustments

B.3.1 Problem description

NTC has the overall responsibility of the railway operations and coordinates the disruption management process. When a disruption is recorded, NTC evaluates its effect and, if it is considered as severe, NTC tries to re-schedule the events of the timetable affected by the disruption.

The severeness of a disruption is not easily assessed. It is described as a combination of how much time will pass until the operations are according to plan again and how many trains will be affected. The number of passengers affected by a disruption also contributes to its degree of severeness. Finally, it makes a large difference to the severeness whether the headways between trains are small or large. For example, the effect caused by a blockage will be less on sections of the network with much time between the trains than on sections with little time between the trains.

Timetables are constructed with included buffer time. Therefore, a timetable is able to absorb some disruptions. Buffer times are included in the dwell times, the running times, and the headways. When a disruption occurs, the buffer times in the timetable are used to gain time whenever possible. Thus they enable recovery from a disruption.

The residual capacity of a railway network at a specific point in time is, as described in Section 2, a concept describing the capacity of the network in operation in relation to the traffic, i.e. how many trains are operated relative to the conditions of the network.

When a severe disruption occurs and it can not be absorbed by the buffers in the timetable, the utilization of the network decreases, and trains may queue up. In that case, NTC aims to increase the residual capacity in the network either by moving trains faster through the network, allowing overtaking at relevant stations, turning trains earlier, canceling departures, etc. Residual capacity is maintained by controlling the traffic flowing in the network and by preventing blocking situations to occur.

In Sections B.3.2 and B.3.3 we distinguish between disruptions with low and high impact on the timetable. Low level impact disruptions are those where recovery to the originally planned timetable is possible by using so-called dispatching rules. High level impact disruptions are those where recovery in this way is not possible, for example if a complete blockage occurs at some part of the network. In such a case, more significant recovery measures are needed.

A survey of optimization models for railway related problems is given by Cordeau et al. [1998]. This survey describes various optimization models developed for railway problems. One of the described problems is the Train Dispatching Problem (TDP). TDP is the problem of minimizing delays by scheduling meets and overtakings, thereby taking into consideration operational costs. The velocity of trains is included in TDP as a decision variable.

Recently, a survey of algorithms and models for railway traffic scheduling and dispatching was given by Törnquist [2006]. The problems mentioned are subdivided into tactical and operational scheduling and re-scheduling. Of specific interest is re-scheduling of trains, which focuses on the re-planning of an existing timetable when a disruption has taken place.

B.3.2 Dispatching rules

Dispatching rules are used on disruptions that have a lower level of impact on the railway system. Dispatching rules are further divided into three subgroups according to the level of severeness of the disruption that invoked them. For disruptions with the lowest level of impact, where no substantial decrease in utilization has yet emerged, it is sufficient to make few modifications to the timetable. At the next level, where the traffic is more affected by the disruption,

it is necessary to increase the utilization of the network. This can be done e.g. by increasing the operational capacity, for example through changes in the timetable in stopping patterns. The severest of the low impact disruptions need an increase in residual network capacity before recovery to a state with larger utilization (corresponding to the original timetable) is possible.

The different rules have different abilities to relieve disruptions and they have different effects for the passengers. From the passengers' point of view, a rule may affect the number of train departures per station or it may force the passengers to change their routes. The effect of a dispatching rule on the delays of trains and its effect on the passengers can be conflicting. Increasing the residual capacity often implies a decrease in the number of train departures, which is undesirable from the passengers' point of view. However, not increasing the residual capacity will make it very hard to absorb a delay, and this is also undesirable for the passengers.

B.3.2.1 Overtaking and changing stopping patterns

Handling operations is less complex if there is a predetermined order of train lines. In the case of a disruption, the predetermined order of lines can be broken on stations with multiple platforms in the same direction i.e. where overtaking between trains is possible. This is, for example, used when a fast train reaches a delayed stop train at a station with two platforms available in the same direction.

If a stop train is delayed and a fast train catches up with it, another possibility is to change their stopping patterns provided that the two trains are of the same rolling stock type and that it is impossible for the fast train to overtake the stop train. This rule is specifically used at S-tog. In practice, the passengers on the stop train are informed that after the next stop their train becomes a fast train. This enables them to get off in time if their destinations are stations where the fast train does not stop. The passengers on the fast train are informed similarly that their train becomes a stop train.

Note that in using both these rules no passenger experiences an additional delay on top of the initial delay caused by the disruption. If no action is taken in the latter situation, the fast train will queue up behind the stop train.

B.3.2.2 Inserting an on-time train

A dispatching rule, which is often used to prevent delays to spread over the network, is the insertion of an on-time train at an intermediate station. If a train is delayed at the first part of its route, it may be possible to insert a replacement train at an intermediate station on the route. The replacement train is inserted according to schedule. When the delayed train reaches the intermediate station, it is taken out of service. Seen from the passenger point of view, fewer departures are delayed. The rule has a limited effect on the overall delay. As no departures are canceled, no residual capacity is created.

B.3.2.3 Increasing Residual Capacity

Residual capacity is increased when departures are canceled. Canceling a departure from a terminal will increase the residual capacity along the entire route of the train. However, from the point of view of NOC, it leaves a train of some composition at the departure terminal. This might also force the cancelation of a departure at the terminal at the other end of the line. It may also create parking capacity problems at the shunting areas.

An alternative to canceling a departure completely is to skip stations along the route of a train, i.e. to change the stopping pattern of the train by decreasing the number of stops along its route. Stops canceled are mostly at stations with minor passenger loads and few connecting lines.

Yet another alternative is to shorten the routes of trains. A train can be turned around before reaching its terminal, i.e. the remaining stations on its route are skipped, cf. Figure C.2. Note that this is a dispatching rule for individual trains, in contrast to the emergency scenarios described in Section B.3.3 where the routes of all trains of a line are shortened temporarily.

Finally, it is possible to cancel an entire train line. An example of how this dispatching rule is used in practice is the cancelation of line B+, which is a line in the present S-tog timetable, cf. Figure B.1. Suppose there is a delay in Hellerup. Due to signaling problems, the trains must run slower than indicated by the timetable. The lines operated on this route are lines A and E running from Hillerød and lines B and B+ running from Holte. To enable better absorption of the ongoing disruption, NTC decides to cancel line B+. The cancelation of line B+ decreases the network utilization thereby allowing an increase in the headways between the remaining trains. In practice, the line is canceled by shunting trains on line B+ to shunting areas as these are reached along the

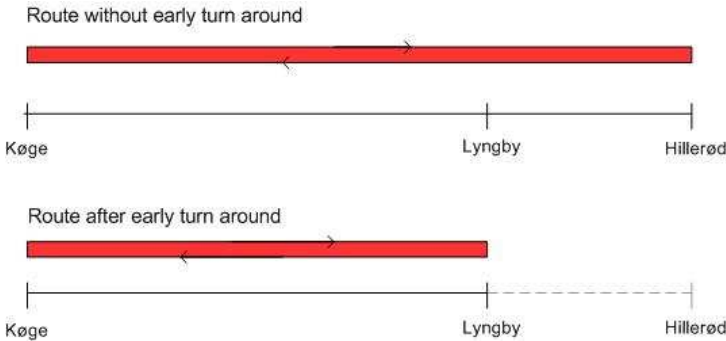


Figure B.5: The train movement at early turn around

route of line B+. Software for planning the later re-insertion of a canceled line is described in Section 4.3.

The advantages of the described dispatching rules are that they all increase the residual capacity for absorbing delays in the disrupted situation. The passengers, however, will experience that there are less departures, which may obstruct their travel plans. Also, if there was no time to couple extra train units to the trains still in operation, the seat capacity of these trains is most likely insufficient. Customer questionnaires show that, like delays and canceled departures, this is also considered as poor quality of service.

B.3.3 Larger disruptions

For high impact disruptions, a set of emergency scenarios may exist, e.g. when tracks in one or both directions are completely blocked. Usually, there is a separate plan for each section in the network and each direction.

The immediate reaction to a high impact disruption is to apply an appropriate emergency scenario. Usually, the headways are so tight that the system will queue up immediately if no adequate measures are taken after a high impact disruption has occurred. Therefore, usually all railway traffic is canceled around the disrupted area. Trains may be turned as closely as possible (according to their usual stopping pattern) to this location. Otherwise, trains may be rerouted, but this requires sufficient capacity on the detour route. Finally, some lines may be canceled completely.

As an example, consider a situation in which the tracks in both directions between stations Dyssegård and Buddinge near Copenhagen are blocked, see Figure B.1. The lines crossing this section in a normal situation are the lines A+, H, and H+. Line A+ is operated between Køge and Buddinge, and lines H and H+ are operated between Frederikssund and Farum. The emergency scenario for this blockage is presented in Tables B.4 and B.5.

Table B.4: Changes of the lines on the section Dyssegård to Buddinge

Line	Changed from and to	Canceled from and to
A+	Køge to Østerport	Østerport to Buddinge
H	Frederikssund to Dyssegård	Dyssegård to Buddinge
	Buddinge to Farum	
H+	Frederikssund to Svanemøllen	Svanemøllen to Farum

Table B.4 shows how the lines are changed and whether they are canceled partly or fully. Unless other disruptions occur, only the lines directly involved in the blockage are included in the emergency scenario.

Table B.5 specifies how many trains are necessary and which turnaround times must be used for them. Each line is changed according to its stopping pattern. Lines A+ and H+ are shortened, and therefore they can be run by 6 and 8 trains, respectively, whereas 8 and 10 trains are necessary normally. Line H is split into two parts and needs 8 plus 3 trains in the disrupted situation, whereas 10 trains are necessary normally.

Given the information in Tables B.4 and B.5, NTC knows which lines to cancel, where to launch bus-services, how many trains to use for each line, and how many train units to shunt to shunting areas.

B.3.4 A comparison with the airline industry

Due to the key differences in infrastructure of the underlying network, disruptions in the airline industry are handled differently than in the railway industry. The air transportation equivalent of NTC is Air Traffic Control, however, one cannot in general view ATC as an infrastructure manager. ATC is responsible for the air traffic with respect to safety both in the airports (airport control), and on the route of an aircraft (en-route control). Another difference is the

Table B.5: Turnaround times and necessary numbers of trains

Line	Traffic south of blockage	Traffic north of blockage
A+	<u>Køge to Østerport</u> Turnaround time: 10 min. Trains necessary: 6	Canceled
H	<u>Frederikssund-Dyssegård</u> Turnaround time: 19 min. Trains necessary: 8	<u>Farum-Buddinge</u> Turnaround time: 13 min. Trains necessary: 3
H+	<u>Frederikssund-Svanemøllen</u> Turnaround time: 16 min. Trains necessary: 8	Canceled

number of operators sharing both airports and the airspace, which is usually much larger than what is experienced in the railway sector.

Disruptions are in some sense much more serious for airlines than for railway companies, because the schedule contains much fewer connections between each origin and destination. Thus, a disruption usually has a much larger impact for the individual airline passengers than for railway passengers. Even then, the general pattern for dealing with a disruption in the airline sector is the same as in the railway sector: First solve the aircraft problem, then the crewing problem, then slots and gates, and then finally the passengers.

When an airline company experiences a disruption, the possibilities regarding timetable changes are very few: Either a departure can be delayed or it can be canceled. In the case of delay, the airline is in the same situation as a railway company: The aviation authorities have to assign a new slot-time, and this requires free slots both in the relevant airports and on the route to be flown. Even though most traffic is routed through corridors in the airspace, the number of possible routes of an aircraft is not bound to a set of tracks layed out in 2 dimensions. From that point of view, the airline problem is much less complex than the corresponding railway problem.

Canceling an aircraft is always possible. However, this is considered to be the worst solution possible. The airlines are normally not bound to a contract specifying the service level and the amount of transportation to be delivered. Instead, competition among airlines servicing routes between the same destinations is a driving force in keeping the service level high.

B.4 Rolling stock re-scheduling

B.4.1 Problem description

This section describes rolling stock re-scheduling in a disrupted situation. Here the assumption is that, whenever this is necessary, the timetable has already been adjusted to the disrupted situation. The main goal is to decide how the rolling stock schedules can be adjusted to this new timetable at reasonable cost and with a minimum amount of passenger inconvenience.

The most characteristic feature of rolling stock is that it is bound to the tracks: rolling stock units cannot overtake one another, except at locations with parallel pairs of tracks. A broken rolling stock unit may entirely block the traffic – actually, this is a frequent cause of disruptions. Moreover, the operational rules of rolling stock units are largely determined by the shunting possibilities at the stations. Unfortunately, shunting is a challenging problem in itself, even for a medium-size station. Therefore, NOC must constantly keep contact with LOC and check whether or not their intended measures can be implemented in practice. The modifications may be impossible due to lack of shunting drivers or infrastructure capacity.

In case of a disruption, the first dispatching task is to keep the railway system running. These first decisions are taken under high time pressure. Timetable services must be provided with rolling stock of any type. Also, the assignment must fulfill some elementary requirements. For example, the rolling stock type must be compatible with the assigned trajectory, and each train should not be longer than the shortest platform on its route. Especially in a disrupted situation, shunting operations are reduced as much as possible. In particular, shunting operations at locations or points in time where they do not occur in the original schedules are highly undesirable.

Railway operators usually keep a certain amount of rolling stock on stand-by. These units can be used only in case of disruptions. Moreover, many of the rolling stock units are idle between the peak hours, since the rolling stock capacity is usually too large for off-peak hours. If a disruption takes place during off-peak hours, these idle units can act as stand-by units.

As a consequence of the first applied measures, the rolling stock units will not finish their daily duties at the locations where they were planned prior to the disruption. This is not a problem if two units of the same type get switched: rolling stock units of the same type can usually take each other's duty for the rest of the day. More likely, however, the numbers of units per type ending up

in the evening at a station differ from the numbers of units per type that were planned to end up there. Thus, unless expensive deadheading trips are used, the traffic on the next day is influenced by the disruption. Modifications of the schedules for the busy peak hours of the next morning are highly undesirable. Therefore additional measures are to be taken so that the rolling stock balance at night is as close to the planned balance as possible. This problem is studied by Maróti [2006].

Like disruption management in general, rolling stock re-scheduling has a stochastic character. For example, it can often only be *estimated* how long it will take to re-open certain temporarily unavailable infrastructure. Also, additional delays are likely to occur in a disrupted situation. Therefore, the dispatchers at NOC and LOC focus on the immediately forthcoming time period only, since planning for a longer period of time may be a waste of effort. They identify possible conflicts, and handle them in order of urgency.

After a disruption, it is preferable for the rolling stock schedules to return to the originally planned schedules as quickly as possible, since the feasibility of the originally planned schedules has been checked in detail.

A further important element in rolling stock re-scheduling is maintenance of rolling stock. Train units need preventive maintenance after a certain number of kilometers or days, roughly once a month. Due to efficiency reasons, units are usually in service just until they reach a certain maintenance limit. Units that are close to this limit and have to undergo a maintenance check in the forthcoming couple of days are monitored permanently. The latter is particularly important during and after a disruption which may have distracted the units from their planned route towards a maintenance facility. NOC has to make sure that these units reach a maintenance facility in time. Usually, only a small number of rolling stock units is involved in planned maintenance routings. Other units of a given type are interchangeable, both in the planning and in the operations.

The airline industry has similar processes when considering the shorthaul part of their operation, however, there are substantial differences for the longhaul part as described in the succeeding section.

B.4.2 Aircraft re-scheduling

The overall goal in airline disruption management is similar to the goal in railway disruption management: to get back to the optimized schedules with causing as little inconvenience for the passengers as possible.

A main difference between airline and railway systems is that trains usually consist of *several* rolling stock units. Moreover, the order of the rolling stock units in the trains may be relevant. Rolling stock units therefore interact in a more complex way with each other than aircraft do.

Moreover, pilots usually have a license for only one or two aircraft types, so swapping aircraft types inevitably leads to large-scale modifications of the crew schedules. As a consequence, the previously assigned aircraft type is changed in re-scheduling only if this is unavoidable. In order to reduce this problem, modern aircraft types may be split into families that can be flown by a single license. If each crew member has a license for just one type, the problem decomposes into subproblems for each fleet type. In a railway context, lack of knowledge about the rolling stock type is much less binding, since most train drivers have licenses for several rolling stock types. Thus rolling stock dispatchers have more freedom to modify rolling stock types.

Another important difference between airline and railway systems is the maintenance strategy. In the airline industry, each aircraft must undergo a larger safety check every 3 to 4 days – this can take place only at a small number of hubs. Therefore maintenance is often taken into account already in early planning phases when creating rotations for individual aircraft.

The term “tail numbering” or “tail assignment” is used for the process of assigning specific aircraft to specific departures. For shorthaul operations this happens only a few days before the day of operation. Therefore, the rotations of aircraft are constructed to be maintenance feasible, i.e. to allow for maintenance checks within the intervals required by the aviation authorities.

For longhaul operations, the maintenance checks are also included in the rotations, but tail numbering takes place earlier than in the shorthaul case. In general, the rotations are planned to allow for some irregularities while maintaining maintenance feasibility.

Railway networks may contain many interconnected train lines. Most rolling stock units serve in a dozen of timetable services every day. This provides more exchange and correction possibilities for rolling stock units than what is usual in

airline cases. A decision on aircraft routing can easily be irrevocable for many hours and in case of a longhaul operation even for a few days.

Finally, from a revenue point of view, cancelation of a train is much less costly than that of a flight.

In the past years, substantial research has been done on aircraft re-scheduling. Kohl et al. [2004] and Clausen et al. [2005] give excellent overviews.

B.4.3 Rolling stock re-scheduling at S-tog

In the case of a disruption affecting the rolling stock schedules, NOC re-allocates rolling stock units to the train tasks. First of all, they aim to cover all tasks sufficiently with respect to the number of seats. There might not be enough time for shunting in each specific case i.e. allocating the right number of train units to a train is not possible. In this case, a train with a seat shortage is preferred over a canceled train.

At some rolling stock depots, space is an issue. Therefore, there can be some difficulties in finding a feasible rolling stock re-allocation.

Positioning data is not automatically supplied to the Rolling Stock Control System (RSCS) at S-tog. The data in the RSCS must therefore be updated manually by NOC during the operations. The updating of data is used respectively for reporting and statistics, and for giving information on the train lengths in real-time to the passengers. Having this information, the passengers will be able to locate themselves correctly on the platforms. As no automatic decision support or optimization system is available, the first feasible solution found is the one implemented in operation.

As mentioned in Section B.3.3, a recovery method employed for large disruptions is canceling train lines. NOC at S-tog has the responsibility of determining a plan for the re-insertion of the train lines after the disruption. A model has been constructed for finding an optimized re-insertion plan, see Jespersen Groth and Clausen [2006]. Based on the given number of trains that must be re-inserted from each depot along the line and the start time of the re-insertion, the model calculates which trains must be re-inserted from which depots, and how the drivers for these trains can get to these depots. The automatic decision support system for re-inserting train lines is used in the operations. Moreover, in an ongoing project, the problem of re-allocating rolling stock units to trains in the operations is addressed.

B.4.4 Rolling stock re-scheduling at NS

A remarkable property of the Dutch railway system is its density. This basically allows for many alternative rolling stock schedules through exchanges of train units. However, usually trains have short turn-around times, which rules out complex shunting operations at end points. Also, the shunting capacity (shunting area and crews) of stations is often a bottleneck.

Another complicating factor is that NS operates rolling stock of different types. Moreover, a train may contain units of different types. In this case, the order of the train units in the train is important. On one hand, this allows adjusting the rolling stock types well to the passenger demand. In case of disruptions, however, the dispatchers have the additional task of monitoring and re-balancing exchanged rolling stock types.

NOC and LOC of NS use an information system for monitoring and adjusting the rolling stock schedules. Tracking and tracing of train units provides information on the real-time locations of individual units. Moreover, the system matches the train units as well as possible with the duties in the actual version of the schedule. Since returning to the original schedule is important, the system represents the actual rolling stock schedule in terms of deviations from the original schedule.

The system does not include optimization modules, it only gives a warning if the rolling stock schedule has time or location conflicts. A new generation decision support system is currently being developed featuring an improved user interface and the possibility to incorporate optimization tools. These optimization tools are developed as part of on-going research at NS. The applicability of the models proposed by Fioole et al. [2006] and Maróti [2006] in the real-time operations will be further explored.

In the Netherlands, maintenance checks on rolling stock units can be carried out only at a few maintenance facilities. Therefore units routed for maintenance are paid special attention in the operations. Maróti and Kroon [2007, 2005] describe two integer programming models for maintenance routing. They take a rolling stock schedule of a few days as input and modify it so that the units that require maintenance soon can reach a maintenance facility in time. The complexity of the problem is analyzed and a heuristic solution approach is suggested and tested on data of NS.

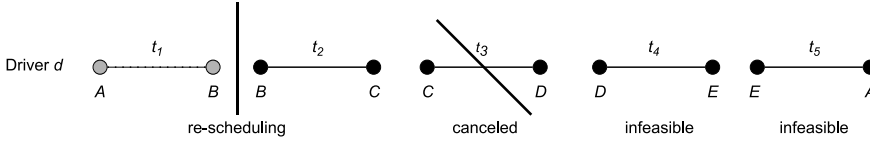


Figure B.6: An infeasible duty

B.5 Crew re-scheduling

B.5.1 Problem description

Recall that the recovery of the timetable, the rolling stock schedule, and the crew schedule is usually done in a sequential fashion. For an estimated duration of the disruption, a modified rolling stock schedule has been determined for a modified timetable. Both are input for the *crew re-scheduling problem*, in which the crew schedule needs to be modified in order to have a driver and an appropriate number of conductors for each task of the modified timetable. Tasks can be either passenger train movements, empty train movements, or shunting activities.

The modified timetable contains the unchanged tasks from the original timetable which have not yet started and additional tasks which were created as reaction to the disrupted situation. For re-scheduling, the set of tasks of the modified timetable can be split into two subsets. The first subset contains all *closed* tasks, which are all tasks that are unchanged, not yet carried out, and part of an original duty which is still *feasible*. The second subset contains the *open* tasks, which include all additional tasks and all unchanged tasks that are assigned to an original duty which has become *infeasible*. A duty becomes infeasible due to a time or a location conflict. The latter may occur, e.g. when one of its tasks has been canceled, and hence the corresponding driver cannot execute the remaining part of his duty.

In Figure B.6, we show an example of an infeasible duty. Because of a disruption, the train containing task t_3 is canceled. Driver d has already finished task t_1 and is at station B . He can perform the next task in his duty, but since t_3 is canceled he cannot go from station C to D . Hence, he will not be able to perform the two last tasks of his duty. Furthermore, this means that, if no action is taken, these two tasks need to be canceled as well. Moreover, driver d has to get back to his crew depot at station A in an appropriate way and at a reasonable time.

In order to prevent additional cancelations due to infeasible duties, the crew re-scheduling problem seeks to assign all open tasks to a crew member. A first possibility that can be used is re-assigning an open task to a crew member of another infeasible duty. Furthermore, an open task may be assigned to a stand-by crew located at a major station.

Since the amount of stand-by crew is limited, a set of feasible duties can also be taken into account for re-scheduling. These duties are broken up and their tasks are added to the set of open tasks. How to determine the set of duties to be broken up is an interesting problem itself. On one hand, the set must be small enough so that the resulting crew re-scheduling problem can be solved quickly, while on the other hand a too small set may not provide enough possibilities to cover the open tasks.

The possibilities for changing duties on the day of operation are based on rules and agreements between the railway company and labor unions. These possibilities usually vary from company to company. For example, the driver's route knowledge has to be taken into account as well as his license for certain rolling stock types. In order to increase the flexibility of the crews, they can be repositioned to another station by traveling on trains as passengers. This option is called crew deadheading.

The objective of the crew re-scheduling problem is a combination of different aspects, namely feasibility, operational costs, and stability. The feasibility aspect is by far the most important, since decisions need to be taken fast in a disrupted situation. It is the decision of the operator how to balance the aspects operational costs and stability.

First of all, there is the feasibility aspect. It is not evident that all open tasks can be covered by a solution. Given two solutions with different uncovered tasks, there may exist a preference for one of them, depending on the urgency and the expected numbers of passengers of the uncovered tasks. If a task cannot be covered, canceling it will lead to a feasible crew re-scheduling solution. An additional cancelation, however, leads to more inconvenience for the passengers, which is against the general aim of disruption management. Moreover, such a cancelation has to be approved by the rolling stock dispatchers and the local planners, since it disturbs the rolling stock circulation. Because a cancelation is a change of the timetable, it has to be approved by NTC.

Operational costs are the second aspect in the objective. In the railway context, the crew payments are often based on fixed salaries. Nevertheless, some parts of a re-scheduling solution influence the operational costs. Crew deadheading on trains can be considered to have no costs other than time, whereas using other transport options for repositioning and taking home stranded crews is

not free. Also, operator specific compensations for extra work due to modified duties need to be considered.

The third aspect in the objective is stability. Humans are involved in the implementation of every re-scheduling solution and can cause its failure. A crew dispatcher may, for example, forget to call a driver and inform him about the modifications in his duty. Therefore, a solution is considered to be more stable if the number of modified duties is smaller.

To the best of our knowledge, only the paper of Walker et al. [2005] deals with re-scheduling of train crews during disruptions. The paper presents a model that manipulates the timetable and the crew schedule at the same time. The objective is to simultaneously minimize the deviation of the new timetable from the original one, and the cost of the crew schedule. One part of the model represents the timetable adjustment, the other part corresponds to a set partitioning model for the crew schedules. Both parts are linked in order to get a compatible solution. It should be mentioned that the railway systems addressed in the research is of a relatively simple structure.

B.5.2 Crew re-scheduling at airlines

Crew re-scheduling has much more effect on the operational cost of an airline operator than of a railway operator. Because of its managerial relevance, airline crew re-scheduling on the day of operation has also become of growing interest for the research community during the last decade.

Yu et al. [2003] reports the savings that Continental Airlines has realized in three major disruptions due to the re-organization of their disruption management process and the installation of decision support systems. The used crew re-scheduling model is based on the prototype described by Song et al. [1998]. A set covering model is formulated, based on a time-space network that represents possible modifications of crew pairings for a certain recovery period. Here a pairing is a sequence of flight legs and overnight rests that begins and ends at the same crew base, and that is to be carried out by a single crew member. This model is solved by depth-first Branch-and-Bound, where open flight legs are covered according to their urgency.

One structural difference between airline and railway crew re-scheduling is the time horizon. Due to more complex regulations for pilots, the position of a pairing within the roster has to be taken into account during re-scheduling (see Medard and Sawhney [2006]). Extending a pairing over the planned duration

can be infeasible due to roster regulations, such as a maximum working time per month. In the railway context, such rules can usually not be violated during re-scheduling. Therefore, usually only duty related rules have to be taken into account for railway crew re-scheduling.

Many approaches in the literature, like Stoiković et al. [1998], Nissen and Haase [2006], Medard and Sawhney [2006], and Lettovský et al. [2000], use column generation to solve set covering or set partitioning models for crew re-scheduling. The first three approaches use network formulations for the subproblems, whereas the last one uses an enumerative pairing generator. We refer to Clausen et al. [2005] for a more detailed description of approaches to airline crew re-scheduling.

B.5.3 Current practice at S-tog

At DSB S-tog a year plan can be changed up to 72 hours before the day of operation, for instance due to work on tracks. Such a plan is called a *special plan*. A very strict restriction in a special plan is the start and end times which can only be moved up to 20 minutes earlier (resp. later). Within the last 72 hours before operations the content of the duties can still be changed without notifying the driver, but the start time cannot be moved earlier and the end time cannot be moved later. If such a move is needed, the planners at NOC must negotiate with the driver.

From 2006 a graphical dispatching system has been used to support the planners. For instance, the drivers have a sign-on terminal and the dispatcher has a real time picture of the drivers meant to sign on during the next half hour. Currently, the system does not contain decision support, which means that all operations are performed manually by planners. The system is currently being extended so that real time information of the train positions are fed to the system. Clearly, without such functionality, it is a tedious process to update the system in major disruptions.

The optimization software, TURNI, described e.g. in Abbink et al. [2005], has been used for generating the annual standard day plans with great success and significant savings during the last couple of years. TURNI is based on a set covering model and dynamic column generation.

Recently, a number of trials have been made to use TURNI also for special plans. The idea used has also been tested at NSR, but due to the smaller problem size at DSB S-tog it seems more likely that S-tog will be able to use TURNI for special planning.

Since the dispatching problem is very similar to the operational planning problem at S-tog, the standard version of TURNI also has been tested for dispatching. The idea is to plan within a window of for instance 2 hours and remove all duties outside the window. The preliminary test with the system shows that approximately 20 minutes is required for a useful solution to be found. Of course, 20 minutes is too much in a disrupted situation, but on the other hand it seems likely that the (exact) solution method is applicable if some time is spent on a more tailored system for dispatching than the standard TURNI system. A potential speedup is to reduce the set of rules from the standard system, since the rules used in dispatching are less restrictive than the rules used for year plans and special plans.

A decision support system for train driver dispatchers is currently under development as a part of a Ph.D.-project supported by S-tog. A solution method to the Train Driver Recovery Problem, described in Rezanova and Ryan [2006], is based on rescheduling a small part of the train driver schedule affected by a disruption. The problem is formulated as a set partitioning problem and possesses strong integer properties. The proposed solution approach is therefore an LP-based Branch & Bound algorithm. The LP-relaxation of the problem is solved with a dynamic column and constraint generation algorithm. Pilot experiments are very promising, both with regards to the integrality property and to the efficiency of the method.

The main objective is to minimize the number of changed duties. The main reason is the resulting communication problem if a large number of duties are changed, since the communication has to be performed manually by the crew dispatcher. A second objective is a robust plan where robustness is defined as large buffer times before breaks within the recovered duties. The main focus in the project is cancelations of entire train series (lines) for a period of time which is commonly used during larger disruptions. This has a large effect on the plans, since many duties are traditionally involved and a p-trip (where the driver travels as passenger) can potentially be canceled making it impossible for the driver to perform his next task.

B.5.4 Crew re-scheduling at NS

The crew dispatchers at NOC of NS use an interactive software system. This provides them with information about the actually planned duties, and enables them to store their duty modifications in the system. The system informs them about delays of trains and about modifications in the timetable and rolling stock schedules. The system also indicates time and location conflicts in the duties.

Recovery options, however, have to be found manually without algorithmic support. In the manual procedure, open tasks are covered one at a time in order of urgency.

Several agreements exist about the way duties may be modified on the day of operation. For example, if a duty is modified, it should not end more than 30 minutes after the end of the original duty. Experiments were carried out to inform crew members automatically via SMS about duty modifications. However, direct communication may be more effective if a dispatcher discovers an option outside the standard rules. Since this negotiation process takes time, the dispatchers often prefer to use stand-by crew to cover open tasks whenever stand-by crew are available.

Recently, Huisman [2005] developed an algorithm for crew re-scheduling in the case of planned track maintenance. The algorithm is based on a combination of column generation and Lagrangian relaxation for solving a set covering type of model. A similar model is used by Nissen and Haase [2006] for airline crew re-scheduling during disruptions. The difference is that, in the case of planned track maintenance, every original duty can be taken into account for re-scheduling, whereas in the latter approach only a subset of the duties is considered due to time limitations.

In an ongoing research project, it will be evaluated if the approach of Huisman [2005] can be adapted to crew re-scheduling during disruptions. The first issue is how to choose the subset of original duties that should be broken-up and taken into account for re-scheduling. Furthermore, acceleration techniques for the column generation process like partial pricing and stabilization will be evaluated. Last but not least, heuristics that produce feasible solutions early in the column generation process may be of great benefit in the context of disruption management.

B.6 Integrated Recovery

In the airline industry the traditional sequence of recovery in case of a disruption is first to resolve the aircraft problem, then to crew this solution, handle the problems regarding infrastructure (gates, arrival/departure slots), and finally to take care of the rerouting of passengers.

This sequence has several drawbacks: Breaking the problem into subproblems may in itself lead to a suboptimal solution of the recovery problem since each subproblem has its own objective. As an example consider a disruption affecting

a short roundtrip from a hub (e.g. Copenhagen - Stockholm - Copenhagen). From a resource point of view canceling the flight is the best reaction since no additional changes to aircraft and crew plans are necessary. However, from a passenger point of view this is the worst solution.

In the past there have been several attempts to construct integrated recovery systems. One approach has been to build dedicated recovery systems for aircraft, crew, and passengers, and then to combine these into an integrated tool. By iterating the recovery process between the dedicated systems this system then tries to find a solution, which from a holistic perspective is better than the individual solutions proposed by each dedicated system. Other architectures have been tried, for example building tools that in one system integrate the recovery of both aircraft and crew, cf. Stojković and Soumis [2001], and approaches taking into account passenger costs cf. Bratu and Barnhart [2004].

Presently no system is capable of true integrated recovery. Due to the development in computational power and in the methods used in dedicated recovery systems, major software vendors as e.g. Jeppesen are, however, optimistic regarding the possibilities of building such system in the airline case.

The situation is quite different in the railway case. Major differences exist regarding the subproblems, which is apparent when one views the processes described in Section 2: In case of a disruption it is the NTC who in the end decides on the solution to be implemented. Furthermore, the possibilities for rerouting passengers are much better - it is often possible to increase the seat capacity of succeeding departures, while this is much more difficult when dealing with aircraft. The integrated recovery approach has therefore received little attention up till now. The benefits from such an approach compared to the sequential approach may, however, be large in terms of quality of service, and the field is expected to become an active research field in the future.

B.7 Conclusions

Railway operators pay much attention to improve their operational performance. One of the key issues is to limit the number of delays by reducing the knock-on effect of single disruptions. To achieve this goal, effective disruption management is required. In this paper, we have explained the role of the different organizations and actors in the disruption management process. An important issue here is that next to the operator itself, the infrastructure manager plays a major role in the disruption management process. The different objectives

of both organizations on one hand and difficult communication schemes on the other hand, complicates the disruption management process a lot.

After the description of disruption management, we have discussed the three subproblems arising in railway disruption management: timetable adjustment, and rolling stock and crew re-scheduling. To adjust the timetable, several different dispatching rules are applied in practice. Unfortunately, no optimization techniques are involved to solve this problem currently. For the re-scheduling of rolling stock and crew some first attempts have been made in the literature to come up with OR models and solution techniques. Most of these have been derived from similar problems in the airline world. However, most of these ideas are in an early stage and have not been applied in practice yet.

In other words, there is a major challenge for the OR community to develop new models and come up with new solution approaches to tackle these problems. Therefore, we hope and expect that another review paper on railway disruption management in about 5 years contains much more models and solution approaches than this one, and moreover that many of them have been applied in practice.

Bibliography

- E. J. Abbink, M. Fischetti, L. G. Kroon, G. Timmer, and M. J. C. M. Vromans. Reinventing Crew Scheduling at Netherlands Railways. *Interfaces*, 35(5):393–401, 2005.
- A. A. Assad. Models for rail transportation. *Transportation Research A*, 14: 205–220, 1980.
- S. Bratu and C. Barnhart. Flight operations recovery: New approaches considering passenger recovery. Working paper, 2004.
- J. Clausen, A. Larsen, and J. Larsen. Disruption management in the airline industry – concepts, models and methods. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2005. URL <http://www2.imm.dtu.dk/pubdb/p.php?3763>.
- J.-F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32:380–404, 1998.
- A. F. de Kort. Advanced railway planning using Max-Plus algebra. In J. Allan, C. A. Brebbia, R. J. Hill, and G. Sciutto, editors, *Computers in Railways VII*, pages 257–266, Ashurst, UK, 2000. WIT Press.
- M. Ehrgott and D. M. Ryan. Constructing Robust Crew Schedules with Bicriteria Optimization. *Journal of Multi-Criteria Decision Analysis*, 11:139–150, 2002.
- P.-J. Fioole, L. G. Kroon, G. Maróti, and A. Schrijver. A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research*, 174:1281–1297, 2006.

- D. Huisman. A Column Generation Approach to solve the Crew Rescheduling Problem. *to appear in European Journal of Operational Research*, 2005.
- D. Huisman, L. G. Kroon, R. M. Lentink, and M. J. C. M. Vromans. Operations Research in passenger railway transportation. *Statistica Neerlandica*, 59(4): 467–497, 2005.
- T. Huisman and R. J. Boucherie. Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B*, 35:271–292, 2001.
- J. Jespersen Groth and J. Clausen. Optimal reinsertion of cancelled train line. Technical Report Report–2006–13, Informatics and Mathematical Modelling, Technical University of Denmark, August 2006.
- N. Kohl, A. Larsen, J. Larsen, A. Ross, and S. Tiourine. Airline disruption management – perspectives, experiences and outlook. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK–2800 Kgs. Lyngby, 2004. URL <http://www2.imm.dtu.dk/pubdb/p.php?3311>.
- L. G. Kroon, R. Dekker, and M. J. C. M. Vromans. Cyclic Railway Timetabling: a Stochastic Optimization Approach. In F. Geraets, L. G. Kroon, D. Wagner, and C. Zaroliagis, editors, *accepted for Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science*, Berlin, 2005. Springer.
- L. Lettovský, E. L. Johnson, and G. L. Nemhauser. Airline Crew Recovery. *Transportation Science*, 34(4):337–348, 2000.
- G. Maróti. *Operations Research Models for Railway Rolling Stock Planning*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2006.
- G. Maróti and L. G. Kroon. Maintenance Routing for Train Units: the Transition Model. *Transportation Science*, 39(4):518–525, 2005.
- G. Maróti and L. G. Kroon. Maintenance Routing for Train Units: the Interchange Model. *Computers & Operations Research*, 34(4):1121–1140, 2007.
- C. P. Medard and N. Sawhney. Airline crew scheduling from planning to operations. *European Journal of Operational Research*, 2006.
- D. Middelkoop and M. Bouwman. Train network simulator for support of network wide planning of infrastructure and timetables. In J. Allan, C. A. Brebbia, R. J. Hill, and G. Sciutto, editors, *Computers in Railways VII*, pages 267–276, Ashurst, UK, 2000. WIT Press.
- R. Nissen and K. Haase. Duty-period-based network model for crew rescheduling in European airlines. *Journal of Scheduling*, 9:255–278, 2006.

- ProRail. poorcapaciteit eerste helft 2006 stabiel t.o.v. 2005 track capacity first half 2006. Internet: <http://www.prorail.nl/ProRail/Bedrijfsinformatie.../Nieuws/Spoorcapaciteiteersteerستهelft2006stabiel+t.o.v+2005.htm>, 2006.
- N. J. Rezanova and D. M. Ryan. Solving the train driver recovery problem. Technical Report Report-2006-24, Informatics and Mathematical Modelling, Technical University of Denmark, December 2006.
- A. J. Schaefer, E. L. Johnson, A. J. Kleywegt, and G. L. Nemhauser. Airline Crew Scheduling Under Uncertainty. *Transportation Science*, 39(4):340-348, August 2005.
- A. Schöbel. Integer programming for solving the delay management problem. Technical Report 92, Technische Universität Kaiserslautern, Institut für Techno- und Wirtschaftsmathematik, January 2004.
- S. Shebalov and D. Klabjan. Robust Airline Crew Pairing: Move-up Crews. *Transportation Science*, 40(3):300-312, 2006.
- M. G. Sohoni, E. L. Johnson, and T. G. Bailey. Operational airline reserve crew planning. *Journal of Scheduling*, 9:203-221, 2006.
- M. Song, G. Wei, and G. Yu. A Decision Support Framework for Crew Management During Airline Irregular Operations. In G. Yu, editor, *Operations Research in the Airline Industry*. Kluwer Academic Publishers, Boston, 1998.
- G. Soto y Koelemeijer, A. R. Iounousov, R. M. P. Goverde, and R. J. van Egmond. PETER, a performance evaluator for railway timetables. In J. Allan, C. A. Brebbia, R. J. Hill, and G. Sciutto, editors, *Computers in Railways VII*, pages 405-414, Ashurst, UK, 2000. WIT Press.
- M. Stoiković, F. Soumis, and J. Desrosiers. The Operational Airline Crew Scheduling Problem. *Transportation Science*, 32(3):232-245, August 1998.
- M. Stojković and F. Soumis. An Optimization Model for the Simultaneous Operational Flight and Pilot Scheduling Problem. *Management Science*, 47: 1290-1305, 2001.
- J. Törnquist. *Railway Traffic Disturbance Management*. PhD thesis, Blekinge Institute of Technology, Karlskrona, Sweden, 2006.
- C. G. Walker, J. N. Snowdon, and D. N. Ryan. Simultaneous disruption recovery of a train timetable and crew roster in real time. *Computers & Operations Research*, 32:2077-2094, 2005.
- J. W. Yen and J. R. Birge. A Stochastic Programming Approach to the Airline Crew Scheduling Problem. *Transportation Science*, 40(6):3-14, February 2006.

G. Yu, M. Argüello, S. Gao, S. M. McCowan, and A. White. A New Era for Crew Recovery at Continental Airlines. *Interfaces*, 33(1):5–22, January–February 2003.

APPENDIX C

**Robustness and Recovery in
Train Scheduling - a
simulation study from DSB
S-tog a/s**

IMM-Technical Report-2006-12. Accepted for publication in International Journal of Operations Research, special issue for Nordic MPS 2006.

Robustness and Recovery in Train Scheduling - a simulation study from DSB S-tog a/s

M. Hofman, L. Madsen, J. J. Groth, J. Clausen, J. Larsen

Department of Informatics and Mathematical Modelling, The Technical University of
Denmark

Abstract

This paper presents a simulation model to study the robustness of timetables of DSB S-tog a/s, the city rail of Copenhagen. Dealing with rush hour scenarios only, the simulation model investigates the effects of disturbances on the S-tog network. Several timetables are analyzed with respect to robustness. Some of these are used in operation and some are generated for the purpose of investigating timetables with specific alternative characteristics.

C.1 Background

DSB S-tog (S-tog) is the sole supplier of rail traffic on the infrastructure of the city-rail network in Copenhagen. S-tog has the responsibility of buying and maintaining trains, ensuring the availability of qualified crew, and setting up plans for departures and arrivals, rolling stock, crew etc. The infrastructural responsibility and the responsibility of safety lie with Banedanmark, which is the company owning the major part of the rail infrastructures in Denmark.

The S-tog network consists of 170 km double tracks and 80 stations. At the most busy time of day the network presently requires 103 trains to cover all lines and departures, including 4 standby units. There are at daily level 1100 departures from end stations and additionally appr. 15.000 departures from intermediate stations. Figure D.1 illustrates the current line structure covering the stations of the network.

All lines of the network have a frequency of 20 minutes and are run according to a cyclic timetable with a cycle of 1 hour. The frequency on stations in specific time periods as e.g. daytime is increased by adding extra lines to the part of

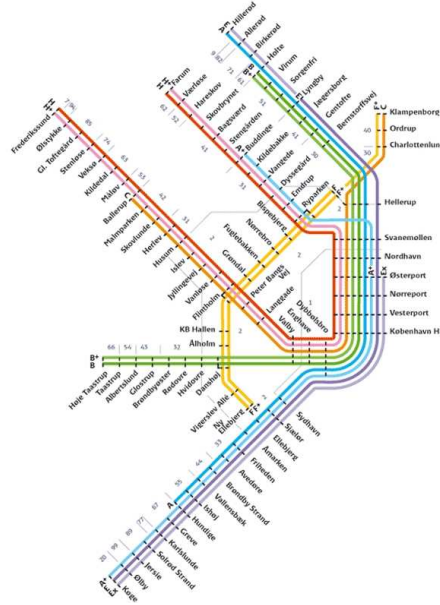


Figure C.1: The DSB S-tog network according to the 2006 timetable

the network covering these specific stations. This way of increasing frequency makes it easy for to customers to remember the line routing both in the regular daytime and in the early and late hours.

Each line must be covered by a certain number of trains according to the length of its route. The trains covering one line forms a circuit. The time of a circuit is the time it takes to go from one terminal to the other and back.

The network consists of two main segments, the small circular rail segment, running from Hellerup in the north to Ny Ellebjerg in the south, and the remaining major network. This consists of seven segments - six "fingers" and a central segment combining the fingers. A consequence of this structure is that a high number of lines pass the central segment resulting in substantial interdependency between these lines. This interdependency makes the network very sensitive to delays and it is thus imperative to S-tog to reduce the line interdependency as much as possible in the early planning stages. The plans of timetable, rolling stock and crew should if possible be robust against disturbances of operations. It is, however, in general non-trivial to achieve such robustness.

C.1.1 Simulation

One way to identify characteristics regarding robustness is by simulating the operation of the network. Simulation helps identifying critical parts of the network, the timetable and the rolling stock and crew plans. One example is poor crew planning in relation to the rolling stock plan. It is unfortunate to have too little slack between two tasks of a driver, if the tasks involve two different sets of rolling stock.

Simulation also provides a convenient way to compare different types of timetables on their ability to maintain reliability in the operation. This allows better decisions to be made on a strategic level regarding which timetable to implement. Specifically, for the network structure of S-tog the number of lines intersecting the central segment has proven important to the stability in operation in the past. It has been a common understanding that an increasing number of lines passing the central segment will lead to a decreasing regularity.

Time slack is often used as a remedy for minor irregularities at the time of operation. Time slack can for example be added to running times along the route, dwell times on intermediate stations and turn around times at terminals. Common for these types of slack are that they are introduced at the time of timetabling in the planning phase.

It is common knowledge that time slack increases the ability of a timetable and a rolling stock plan to cope with the facts of reality, i.e. the unavoidable disturbances arising in operation. Slack in a plan is, however, costly since resources are idle in the slack time if no disturbance occurs. It is therefore not evident which type of slack to use, exactly where to use it, and how much to use.

The stability of a network is not only related to the "inner robustness" introduced through time slack. As noted earlier, slacks in the plans are intended to compensate for minor disturbances. When larger disturbances occur action must be taken to bring the plan back to normal. This process is called recovery. There are various types of recovering plans. For example, cancelling departures decreases the frequency of trains on stations, which in turn increases freedom in handling the disturbance.

The simulation model to be presented is used for testing various timetables with different characteristics. Also we use the model for testing some of the strategies of recovery used by rolling stock dispatchers at S-tog. Firstly, in Section 2, related literature on the subject is presented. Recovery strategies employed at S-tog are described in Section 3. In Section 4 we present the background for the simulation model, and Section 5 discusses assumptions and concepts of the

model. The model itself is presented in Section 6, and the test setups and results are presented in sections 7 and 8. Finally, Section 9 gives our conclusions and suggestions for further work.

More details on the topic can be found in the M.Sc. thesis, Hofman and Madsen [2005].

C.2 Related work

Related work involves studies on robustness and reliability, simulation and recovery. The first subject area, robustness and reliability, focuses on identifying and quantifying robustness and reliability of plans. Simulation is used for various purposes within the rail industry, and the models of the various subjects often have similar characteristics. The area of recovery presents various strategies and systems for recovery. Systems are often based on optimization models.

C.2.1 Robustness and reliability studies

Analytical and simulation methods for evaluating stability are often too complex or computationally extremely demanding. The most common method is therefore using heuristic measures. In Carey and Carville [2000] is described various heuristic measures of stability that can be employed at early planning stages. Carey and Carville [2004] present a simulation model used for testing schedule performance regarding the probability distribution of so-called secondary delays (knock-on effects) caused by the primary delays, given the occurrence of these and a schedule. The model is used for evaluating schedules with respect to the ability to absorb delays. In Vroman et al. is presented concepts of reliability in public railway systems. Using simulation they test the effect of homogenizing lines and number of stops in timetables. Mattson presents a literature study on how secondary delays are related to the amount of primary delay and the capacity utilization of the rail network. An analytic tool for evaluating timetable performance in a deterministic setting, PETER, is presented by Goverde and Odijk [2002]. The evaluation of timetables is done without simulation, which (in contrast to simulation based methods) makes PETER suitable for quick evaluations.

C.2.2 Simulation studies

Hoogheimestra and Teunisse [1998] presents a prototype of a simulator used for robustness study of timetables for the Dutch railway network. The simulation prototype is called the DONS-simulator and is used for generating timetables. Similarly, in Middelkoop and Bouwman [2001] is presented a simulation model, Simone, for analyzing timetable robustness. The model simulates a complete network and is used to identify bottlenecks. Sandblad et al. [2003] offer a general introduction to simulation of train traffic. A simulation system is discussed with the multiple purposes of improving methods for train traffic planning, experimenting with developing new systems, and training of operators.

C.2.3 Recovery studies

In Goodman and Takagi [2004] computerized systems for recovery and various criteria for evaluating recovery are discussed. In particular, they present two main methods of implementing recovery strategies: Either recovering from a known set of recovery rules or optimizing the individual situation, i.e. determining the optimal recovery strategy for the specific instance at hand. A train holding model is presented in Puong and Wilson [2004]. The objective of the model is to minimize the effect of minor disturbances by levelling the distance between trains by holding them at certain times and places of the network. In Kawakami is described the future framework of a traffic control system for a network of magnetically levitated high speed trains in Japan. Different recovery strategies are presented, one of which is increasing the speed of delayed trains.

C.3 Recovery strategies

When a timetable is exposed to disturbances and disruption occurs, it is crucial how the operation returns to normal, and how fast the strategy can be implemented. At present, the procedure of returning to a normal state of operation is manual with support from operation surveillance systems and a system showing the plan of operation constructed in advance. The different manual actions available are mainly the following:

Platform changes on-the-day It is planned in advance which platforms to use for the different train arrivals and departures at the time of operation.

If a planned platform is occupied at the time of arrival of the next train, the train is rescheduled to another vacant platform if possible. For example, at Copenhagen Central (KH) there are two platforms in each direction. When one platform is occupied with a delayed train the trains can be lead to the other vacant platform for that direction.

Trains skipping stations i.e. making fast-trains out of stop-trains If a train is delayed it is possible to skip some of its stops at stations with minor passenger loads and few connecting lines. However, two consecutive departures on the same line cannot be skipped.

Shortening the routes of trains A train can be "turned around" before reaching its terminal i.e. the remainder of the stations on its route can be skipped, cf. Figure C.2. Again, two consecutive trains cannot be turned.

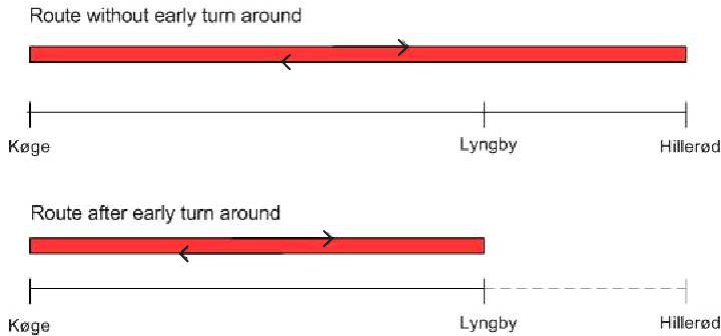


Figure C.2: The train movement at early turn around

Swapping the tasks/routes of fast-trains catching up with stop-trains

On some of the segments of the network both slow trains stopping at all stations and faster trains that skip certain stations are running. Delays some times occur so that fast lines catch up with slow lines leading to a delay of the fast trains. Here, it is possible do a "virtual overtaking", i.e. to swap the identity of the two trains so that the slow train is changed to a fast train and vice versa.

Inserting replacement trains from KH for trains that are delayed Trains covering lines that intersect the central section run from one end of the network to the other passing Copenhagen Central. Here, a major rolling stock depot as well as a crew depot is located. If a train is delayed in the first part of its route, it is often replaced by another train departing on-time from KH. Thus, a new train is set in operation at KH, which

proceeds on the route of the delayed train. This is on arrival at KH taken out of operation.

Inserting replacement trains for trains that have broken down In case of rolling stock failure the train is replaced by new unit of rolling stock from a nearby depot.

Reducing dwell times to a minimum At stations there are pre-decided dwell times. These vary with the different passenger flows of the stations and with different special characteristics such as a driver depot. The latter demands extra time for the releasing of drivers. In the case of a disruption the dwell times on all stations are reduced to minimum.

Reducing headways to a minimum In the outer ends of the network there are some slack on the headways. In the case of delays headways are reduced making the trains drive closer to each other. As the frequency of trains in the central section is high there is less slack here for decreasing headways.

Reducing running times to a minimum Timetables are constructed given predefined running times between all sets of adjacent stations. The running time is always the minimum running time plus some slack. In case of a disruption, running times between all stations are reduced to a minimum given the particular context.

Allowing overtaking on stations with available tracks Handling operations is less complex if there is a predetermined order of train lines. In the case of a disruption the predetermined order of lines can be broken on stations with several available platforms in the same direction i.e. where overtaking between trains is possible. This is for example used when a fast train reaches a delayed stop train at KH.

Cancelling of entire train lines In the case of severe disruption entire lines are taken out, i.e. all trains currently servicing the departures on the relevant lines are taken out of operation. In the case of severe weather conditions such as heavy snow, the decision is taken prior to the start of the operation.

The main components in recovery strategies are increasing headways or exploiting slack in the network, called respectively re-establishing and re-scheduling. The first handles disturbances by employing pre-scheduled buffers in the plans. The latter refers to the handling of disturbances by making some changes in the plan to bring the situation back to normal. The ways of changing the plan are in most cases predefined.

C.4 Background of the problem

C.4.1 Planning and designing timetables

In S-tog the first phase of timetabling consists of deciding the overall line-structure of the train network. The basis for the decision includes various criteria such as number of passengers on the different fingers, passenger travel-patterns and rotation time of lines. Regarding the latter criteria, it is from a crewing perspective an advantage to keep the rotation time at a level matching a reasonable duration for driver-tasks. In the next phase the stopping patterns are decided automatically from input such as driving time, minimum headways and turn-around times. In the third phase, we then verify whether the plan is feasible with respect to rolling stock. These first three first phases are all carried out internally in S-tog. The following phases involve various other parties, each of which evaluates the proposed timetable, including BaneDanmark and the National Rail Authority. When all involved parties have accepted the timetable, the phase of rolling stock planning begins.

The process of designing and constructing a timetable is exceedingly long. It is made up by the long process of constructing possible timetables that might be rejected in other phases of the process, thereby forcing the process of timetabling to be highly iterative. Many stakeholders are involved in the decision of which timetable to implement in operation, and these may very well have conflicting interests. In all phases of the timetabling process there is an urgent need for being able to discuss specific plans both qualitatively and quantitatively. Quantitative information can be obtained by simulation. Often it is an advantage not to have too many details in the input of a simulation. To compare different timetables it may e.g. not be necessary to know all details about tracks and signals. Therefore, a decision regarding the timetable to be developed for operation may be taken early in the planning process.

C.4.2 Disturbances at S-tog

The disturbances at S-tog can be classified into categories at several levels leading to various actions when experienced during operations. First of all, disturbances are categorized as being the consequence of some specific primary incident as e.g. rolling stock defects (causing speed reductions), passenger's questions to the train driver, illness of a driver, or signal problems (forcing the trains to stop). We distinguish between primary incidents caused by the rail system (trains, rails, passengers etc.) and driver related incidents.

Incidents with a very long duration and complete breakdowns of the system are considered as a separate type of incidents. An example of a complete breakdown is the fall-down of overhead wires.

Secondary incidents occur as a consequence of primary incidents. These incidents occur because primary incidents have influenced the operation, forcing trains to stop or to slow down. The slack present in the timetable and the number of secondary incidents that usually occur during operation are directly related. That is, when slack is decreased the number of secondary delays increases and vice versa.

The general measures of disturbances in the S-tog network are termed regularity and reliability. These refer respectively to lateness and cancellations in the network. Regularity is calculated as

$$\left(1 - \frac{LateDepartures}{DeparturesinTotal}\right) * 100\%$$

Traffic is considered stable when regularity exceeds a limit of 95%. A departure is late when it is delayed more than 2.5 minutes. Reliability is calculated as

$$\left(\frac{ActualDepartures}{ScheduledDepartures}\right) * 100\%$$

Contractually, reliability must be higher than 97% over the day.

C.4.3 Recovery strategies

Implementing different recovery strategies in a simulation model makes it possible to evaluate, which actions lead to the quickest recovery and least sizeable disruption with respect to affected trains. We have chosen to investigate three specific S-tog strategies for recovery. These have been implemented in the simulation model and are evaluated individually i.e. two different recovery strategies are not employed at the same time in any of the presented test-cases. The three recovery strategies chosen were "Early turn around", "Insertion of on-time trains on KH" and "Cancelling of entire train lines". All of these recovery strategies are frequently used in operation. They each contribute to increased headways in some segment of the network. Furthermore, these three methods of recovery

are employed both in case of smaller and of medium size delays. Also they have varying effects on customer service level.

Early turn around increases headways in the part of the network not serviced because of the early turn around, and the train catches up on schedule in the following departures. As a result, the number of secondary delays is decreased as the train is often turned to an on-time departure. The negative consequences of the recovery strategy are that some departures are cancelled when the train is turned around before the end station of its route. This decreases the reliability. Also, it becomes difficult to locate the rolling stock according to the circular schedule, which must continue the following morning. In reality the trains are turned without any respect of the line of the train. The train simply turns and departs according to the first scheduled departure.

In the simulation model the strategy has been implemented with the constraint that two successive trains can not be turned, i.e. one of them must continue to the end station to meet passenger demands. Also, a train can not be turned in both ends of its route. The shortening of routes are, apart from these two constraints, invoked for each individual train by judging whether it is either more late than a certain threshold or more late than can be gained by using the buffer at the end station. In principle, it is physically possible to turn around trains on all stations in the S-tog network. However, as only a subset of the larger stations are used for turn around in practice, these are also the only stations in the simulation model where turn around is feasible. In the model, a turned around train must match the departures that was originally planned for that particular train.

Cancelling of entire train lines is invoked by the condition of the regularity of the line in question. If the regularity of the line is below a certain threshold, the line or a predefined extra line on the same route is taken out. The line may be reinserted when the regularity again exceeds a certain lower limit and has been above this limit for a predefined amount of time. When put into action this recovery strategy increases the headways on the segment of the network where the line in question runs. A positive effect of the recovery strategy is that the number of secondary delays decreases. As entire lines are cancelled, employing this strategy has a considerable negative impact on the reliability.

Specific characteristics of the recovery strategy are that trains on the line in question can only be taken out at rolling stock depots and that at the time of insertion it must be ensured that drivers are available at these depots. As drivers are not simulated in the model, the latter restriction is not included.

Insertion of on-time trains on KH is the strategy of replacing a late train with train being on-time from KH. This means that the time the network is serviced

by the delayed train is decreased. Like the recovery strategy of shortening routes, this strategy is also employed when the relevant train is more than a predefined threshold late. The threshold limit is set by the duration of the buffer at end station. The strategy has no impact on the reliability as no trains are being cancelled. It does, though, have a limited positive effect on the regularity. As no headways are increased the headways are merely levelled out in the part of the route from KH to the end station. It is assumed in the model that only one train in each direction on the same line can be replaced at the same time. Hence, at least every second train services the entire line.

C.5 Assumptions

One of the difficulties in simulation modelling is to decide on the level of detail to use, i.e. to decide whether it is necessary to implement a very detailed model or whether trustworthy conclusions can be made on the basis of more coarse grained information. In the rail universe we have to determine whether signals and tracks must be modelled with high precision or whether it is sufficient to model a network with stations as the nodes and tracks between them as the edges.

Additional considerations regarding specific details must also be made. Below we describe the assumptions we have made in modelling the S-tog network.

All experiments are based on the worst case scenario of operating peak hour capacity throughout the simulation. This will not affect the validity of the results as stability and robustness are lowest when production and demand are highest.

We assume that the stopping pattern of each lines is constant over the day. In most cases, each line has a fixed individual stopping pattern over the day. Deviations do occur, especially in the early morning hours and in the evening. As we have chosen only to simulate peak hours not intersecting these time intervals, we assume that the stopping pattern for each line is fixed.

The stopping times of trains in the timetable are given with the accuracy of half a minute. Therefore, the train in reality arrives at a station approximately at the time defined by the timetable. Arrivals "before schedule" may thus occur. Since we do not allow a train to depart earlier than scheduled, these early arrivals have not been implemented in our simulation-model.

The circular rail segment has been omitted from the test scenarios. In general, it

has a very high regularity and its interaction with the remainder of the network is very limited.

In the model, all minimum headways have been set to 1.5 minutes. This makes the model less exact than if minimum headways are kept at their real levels, which vary depending on the area of the network. In reality, network parts where trains drive with high speed have larger minimum headways than low speed parts. However, due to the heavy traffic the low speed parts constitute the bottleneck network parts.

In our model delays are added at stations. The alternative is to add delays between stations describing the track segment between two stations to some predefined detail. This, however, complicates the model without giving any additional benefits regarding the possible comparisons between time tables and recovery strategies.

Delays are generated from delay-distributions of historical data. We hence assume that the delays in the system will occur mainly caused by the same events as they have done up till now. However, there may be a variation in delay patterns stemming from the structure of the timetable. Even if no timetable similar to the timetable in a test scenario have been in operation, the delays observed at stations in the past still seem to offer the best basis for generating delays for the test scenario in question.

The probability of delay on a station is set to 50%. This is estimated from the historical data as a worst case situation. Almost no time registrations are zero (i.e. the departure is exactly on time).

In our model, regaining time is only possible at stations and terminals and not while running between stations. Even though time can be gained between the stations in the outer part of the network, this is insignificant compared to what can be gained in the terminals. Again, it is clear that the regularity of a test case in real-life will be at least as good as the one observed in the simulation model, since extra possibilities for regaining lost time are present.

The single track of 500 m on a part between Værløse and Farum is not modelled. This is the only part of the network with a single track. As the single track part only accounts for 0.3% of the network this has no measurable effect on the results.

In the central section there are four junctions in the form of stations where lines merge and split up. To enable the use of a simple common station model, these junctions are not explicitly modelled in the simulation model. To compensate for this, virtual stations are introduced in the model. On the hub stations,

where different sections of the network intersect, a station is added for merging or parting of the lines meeting at the hub. As a result of the extra station, the model merges and divides at slightly other times than in reality. An example of this is Svanemøllen (SAM). At SAM the northbound track divides into two. Hence, the lines that have passed the central section divide into two subsets. In the 2003 timetable, the subsets are two lines running towards Ryparken (RYT) and the remainder running towards Hellerup (HL). SAM is modelled as four stations; two stations where trains run towards respectively come from RYT and two that run towards respectively come from HL. Going south this means that when departing from SAM the trains must merge so no "crash" appears. When a station has several platforms in each direction, this is also handled in the model by adding in an extra station for each platform. For example, KH is modelled as four stations, two in each direction. This means that KH has two platforms available for each direction and can have up to four trains in the station at the same time.

The changes in the infrastructure since 2003 mostly concern the expansion of the circular rail of the network. Therefore, results obtained using the 2003 structure are still valid.

The simulation model is in general coarse grained and contains several minor modifications in relation to the facts of reality. Nevertheless, the model is adequate for comparing timetables and for evaluating the immediate impact of one recovery method compared to either one of the two other implemented recovery methods or no recovery cf. the text sections above.

C.6 The simulation model

The simulation model has been implemented in Arena Kelton et al. [2004], which is a general programming tool for implementing simulation models. The model is based on the circulations of rolling stock for each of the lines. Therefore, the main model of the simulation is built based on the lines. It has an entrance for each line where entities are created corresponding to the trains necessary to run the line. The trains circulate in a general station submodel common for all stations. A recovery method is given before the entities enter the station submodel and start iterating over it.

The input to the model is the line sequences, the departures, and various station information such as for example whether a particular station is a terminal, an intermediate stopping station or an intermediate non-stopping station, and the dwelling time at each station.

C.6.1 Station submodel

In the station submodel attributes are first updated for the next step and the next station respectively as these are used in the model relative to the current step and station. The model iterates over the stations in each line of the network. Therefore, the model reiterates from the beginning when the final station in the route is reached. Secondly, the attribute of direction is updated depending on the arriving train entity. Thirdly, the entity is put on hold if the station of the current step is occupied by another train. If the station is not occupied, the entity in question is allowed to enter the station. This is emphasized in the model by setting an "occupied" flag on the station. Thereafter, it is decided which type of station is entered, given the three possibilities.

The next action of the station submodel is handling the train dwelling time depending on the type of the station. If the train entity is set to stop at the station, the train is delayed by the predefined dwelling time. The dwelling time assigned depends on whether the train entity is already delayed from a previous station. If the train is delayed it should use the minimum dwelling time allowed. If not, it should use the standard dwelling time. No train can leave earlier than scheduled.

Next a possible delay is added. Delay is added at 50% of the stations. There are no delays added in the model before all trains have been introduced. Delays are added to the trains according to a distribution based on historical data.

The station is now marked unoccupied, as the train leaves the station after have performed its stop including dwelling time and possible delay. The regularity and the reliability are updated immediately after the station has been registered as unoccupied. These are calculated for each train on each of its stations. The overall regularity and reliability are the final averages of the individual values.

Now the entity enters some recovery method depending on which method was chosen initially. The method may be that no recovery action should be taken at all.

After recovery, the specific case of merging the lines B and B+ is handled in the submodel merge. If the line of the train entity is either the B or the B+ line and the current station is Høje Taastrup (HTAA), the trains merge and drive alternately B and B+ unless recovery has cancelled line B+. The merge is handled simply by alternating an attribute on the entity characterizing which line the train entity runs. If B+ has been cancelled, merging is not possible and the trains are instead delayed 10 minutes, which is the frequency between B and B+.

Routing is also handled in the station submodel. In the routing part, the train entity is routed from the current station to the next. First the train is held back to ensure sufficient headway. Next the train is held back in a queue until there is an open platform at the following station. There is a maximum number on the queue length identical to the space on tracks between stations in the S-tog network. If the current station is a terminal, the train can gain time and is routed to the same station in opposite direction otherwise it is routed to the next station in its line sequence without the possibility of gaining lost time. Finally, time is updated for the train entity with the driving from one station to the next.

C.6.2 Recovery submodels

C.6.2.1 Early turn-around

The basic idea of this recovery method is that if a train is delayed more than a certain threshold, it will change direction at an intermediate station before it reaches the planned next terminal. This is checked in the beginning of the model together with a check of whether the line has been turned on its previous trip in the opposite direction.

If the current station is a possible turn-around station, the turn-around is performed and the next step and the starting time are decided. By creating a duplicate of the train entity turned around, it is possible to ensure that the following train is not also turned early.

C.6.2.2 Take Out

This recovery method cancels specific lines in the network in case of disruption. The cancellation of lines are initiated by regularity falling below a certain threshold. When regularity has reattained another certain threshold, the method reinserts the trains on the cancelled line.

The candidates to be cancelled are predefined. For example, if delays are on line A, line A+ is cancelled.

Trains can only be taken out on depot stations. We assume the availability of drivers at the time of reinsertion. The method sets the train entities on hold. The cancellation of some entity is simply done by setting the train entities to

be cancelled on hold and reinsertion is initiated by signalling. Time and station are then updated according to the time on hold and the line of the entity, and the train entity continues to run from that specific station along its planned line sequence.

C.6.2.3 Replace

This recovery method inserts an on-time train from KH to replace a train delayed along its route, which is then taken out. It is activated when a train is more late than a certain threshold and the previous train was allowed to continue along its entire route.

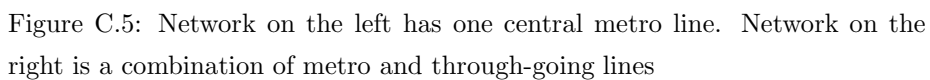
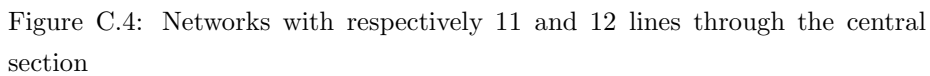
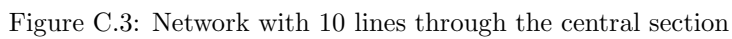
The model of the method is divided in two. One handling the take out of trains at KH and one handling observation of delay at all other stations and scheduled insertion on KH. In the latter of these, a duplicate of the train entity is created to ensure that the train is taken out when it reaches KH.

It is at all times assumed that rolling stock is available at KH for inserting trains.

C.7 Test Cases

For the purpose of testing the simulation model 7 timetables has been used, some of which are run in several versions to make results more comparable. Two of the timetables are actual timetables of respectively 2003 with 10 lines intersecting the central section and 2006 with 9 lines intersecting the central section. They are both of the structure seen in Figure D.1 Three timetables are potential timetables for years to come. They have respectively 10, 11 and 12 lines intersecting the central section. See Figure C.3 and Figure C.4. Finally, two artificial timetables have been constructed especially for the test session. The first of these has 19 lines on the fingers and 1 central metro line in the central section. The other has in total 17 lines, with a combination of circular and drive through lines in the central section. See Figure C.5.

The purpose of the test session with so different timetables is to test the effect of different characteristics such as a varied number of lines, different stopping patterns, line structures, cycle times, homogeneous use of double tracks, homogeneous scheduled headways and buffer times at terminals.



To make results comparable, changes have been made to some of the timetables. For example, lines have been extended and headways have been evened out.

The recovery methods have been tested with varying thresholds for activation of the methods. The *Early Turn around* and *Replace* methods have been tested for activation when the train in question is more late than respectively 2.5 minutes, 5 minutes, and “the amount of buffer time” at the terminal. For the *Cancellation* method, activation has been set at regularity falling below 80% without reinsertion, or 90% both with or without reinsertion. Reinsertion takes place when regularity increases above 95%. The recovery methods are not tested on the artificial timetables as these are so different from the timetables of today that recovery results are incomparable.

A series of tests were run with varying buffer time at terminals.

Tests with small and large delays are performed. In these test cases we have added respectively small delays, large delays and both large and small delays. The definition of small and large delays are derived from the historical data. The delays divide the stations into two subset of respectively 80 stations with small delays and 81 stations with large delays. For the first two of the three test scenarios, delay can hence only occur occur at 50 % of the stations. The tests are run with no recovery and 100% probability of delay on the relevant stations.

C.8 Computational Results

A variety of tests have been carried out with the simulation model. We have chosen to present specifically test results regarding *the comparison of timetables*, *the effect of large versus small delays on operation* and *varying sizes of terminal buffer times*. The complete set of tests is described in Hofman and Madsen [2005].

The main measures used for evaluating results are *regularity* and *reliability*. The registration in the simulation model starts when the start-up period is completed, i.e. when all trains has been inserted in the current model run.

When evaluating the results, it is also interesting to evaluate the cost of a timetable with respect to the number of trains necessary to maintain circulation. An optimal solution is a robust timetable operated by as few trains as possible. This is an obvious trade-off since fewer trains in a solution implies that the times of circuits for lines are decreased. The result is less “room” for slack in the timetable and therefore generally less robustness.

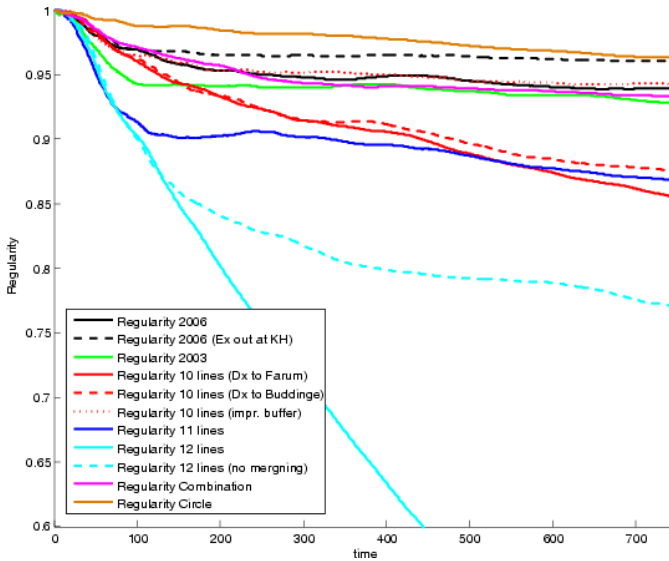


Figure C.6: Regularity of the 12 tested timetables where no recovery is applied

C.8.1 Comparing Timetables without recovery

A total of 12 different timetables has been tested with and without recovery. Figure C.6 shows a plot of the regularity of different timetables run without recovery.

In general the number of lines have a high impact on regularity. Fewer lines implies an increase in regularity. It is, however, possible to improve timetables that has a high number of lines by increasing buffers on terminals. The results show that increased buffers improve the ability to “cope with” delays. An example of this is the timetable with 10 lines, cf. Figure C.3.

C.8.2 Comparing Timetables using Turn-Around Recovery

The regularities of the timetables run with the turn-around recovery method are shown in Figure C.7. The threshold for invoking the method has been set to the terminal buffer time used in the time tables.

Results show again that the number of lines significantly influences the level of regularity, however, the effect decreases with increasing number of lines. This is a consequence of more trains reaching the threshold and hence being turned, cf. Figure C.8, where regularities of timetables are shown with a threshold for the turn-around recovery set to 5 minutes. The ranking of timetables with respect to level of regularity is here different from that of Figure C.7. In addition, an overall better regularity on lines when using buffertimes as threshold can be observed.

C.8.3 Comparing Timetables using Cancellation of Lines Recovery

As expected, the results show that the cancellation of lines has a very positive effect on regularity. Corresponding to the positive effect on regularity, the recovery method has a negative effect on reliability. That is, the majority of departures may be on time but only when a substantial part of the planned departures have been cancelled. The results for all timetables are given in Figure C.9.

C.8.4 Comparing Timetables using Replacement of Trains Recovery

This recovery method does not cancel any departures. Therefore the reliability is 100% in all test results. This also means that the headways are not increased when the recovery method is invoked. As expected this shows that the positive effect on regularity is less than for the other recovery methods.

C.8.5 Comparing the Effectiveness of Recovery Methods

If we compare the results of the “turn-around” with the “line-cancellation” recovery method, we see that the regularity of the “turn-around” is at the same level as the one of “line-cancellation” for timetables with a low number of lines. For timetables with high numbers of lines, only “line-cancellation” recovery brings up the regularity to a sufficiently high level.

Comparing recovery by replacement with the two other recovery methods, it is

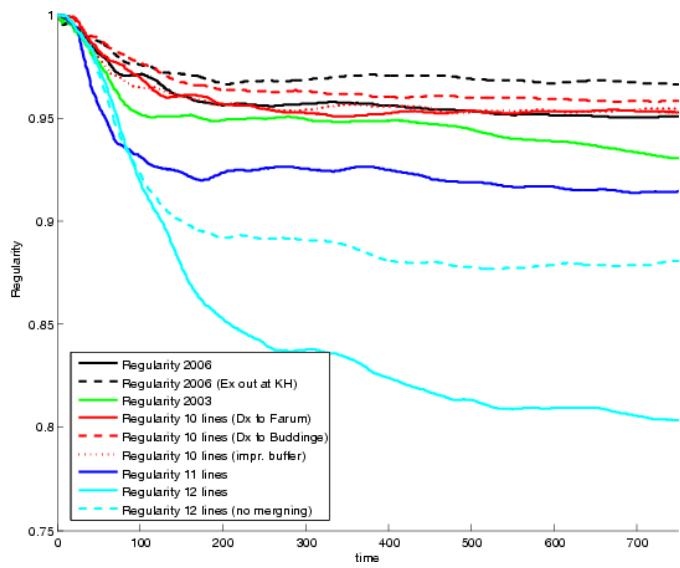


Figure C.7: Regularity of the 12 tested timetables where Turn Around recovery is applied

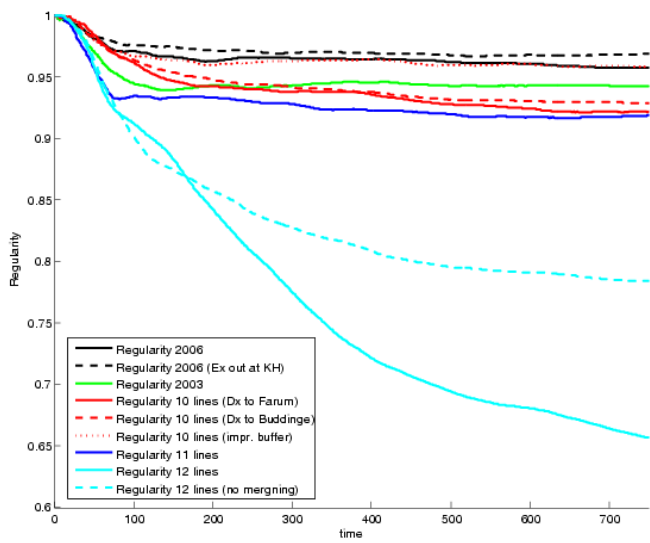


Figure C.8: Regularity of the 12 tested timetables where Turn Around recovery is applied when delay is higher than 5 minutes

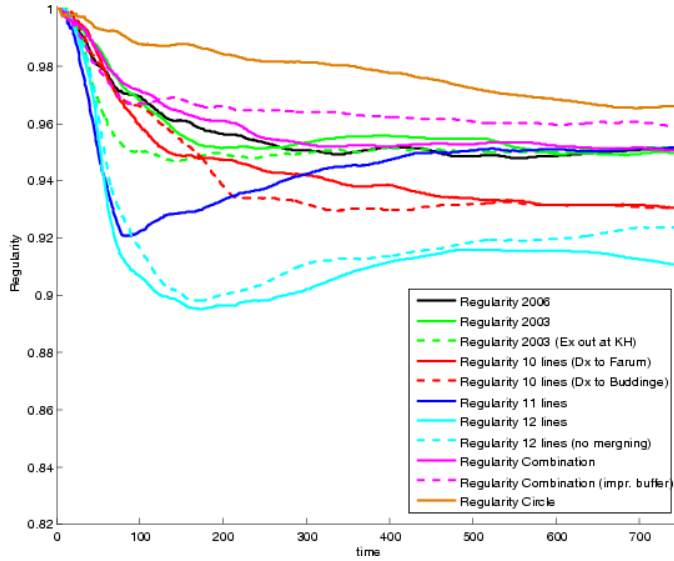


Figure C.9: Regularity of the 12 tested timetables where Cancellation recovery is applied when regularity is under 90%

evident that the method does not have the same level of effect on the regularity as the two others when it comes to the timetables with many lines.

C.8.6 Testing the Effect of Large and Small Delays

The test results of running with small and large delays separately are shown in Figure C.10 for timetables with 12 lines. Similar results were observed for other timetables.

The figure shows a clear tendency: Small delays have almost no effect on the regularity when no large delays are present. The size of buffers are relatively large compared to the delays in the system. Large delays have a significant effect on the regularity as expected. When small delays are introduced in addition to the large delays, they have a much larger effect on propagation of delay than when they occur on their own. It is, however, still obvious that larger delays has the largest effect on regularity and that these if possible should be eliminated. Nevertheless, a substantial increase in regularity can be achieved through the removal of small delays, which is a much easier task.

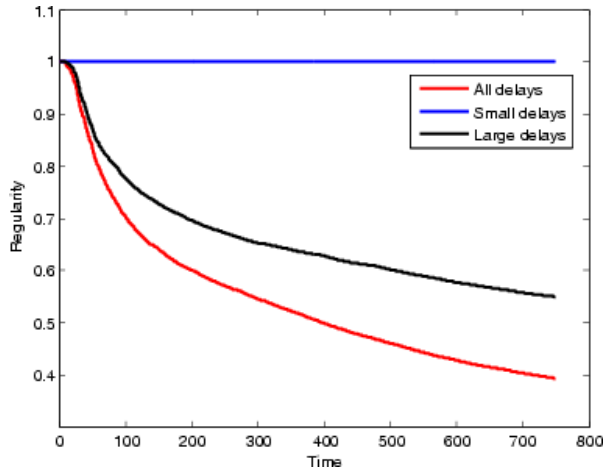


Figure C.10: Regularity when respectively only small delays, only large delays and all delays are applied

C.8.7 Terminal Buffers

The terminal buffers has a substantial effect on regularity. There is often more available time at end stations than on intermediate stations with respect to the size of buffers. As buffers are larger on terminals, there is a better possibility to decrease an already incurred delay. Regarding the size of terminal buffers it is expected that increasing buffer times at terminals in general implies decreasing delays in the network. Tests were run with increasing buffer times to confirm this. The increase in buffer time necessitates that one additional train is set into rotation on specific lines. Hence the number of trains necessary to cover the line increases as the buffers on terminals are increased, cf. Table C.1.

The results show that in general regularity improves when buffers are increased, but also that there is an upper limit on the amount of buffer time, beyond which no extra regularity is gained, cf. Figure C.11 and C.12.

The improvement of regularity depends heavily on the timetable in question for each individual test. The timetable with 12 lines improves considerably more than the timetable with 9 lines.

Timetable	Trains Needed
2003, 10 lines	73
2003, 10 lines and improved buffers on terminals	77
Constructed, 10 lines	67
Constructed, 10 lines and improved buffers on terminals	71
Constructed, 12 lines	93
Constructed, 12 lines and improved buffers on terminals	100
Combination	82
Combination, Improved buffers on terminals	88

Table C.1: Number of trains running simultaneously in the tested timetables

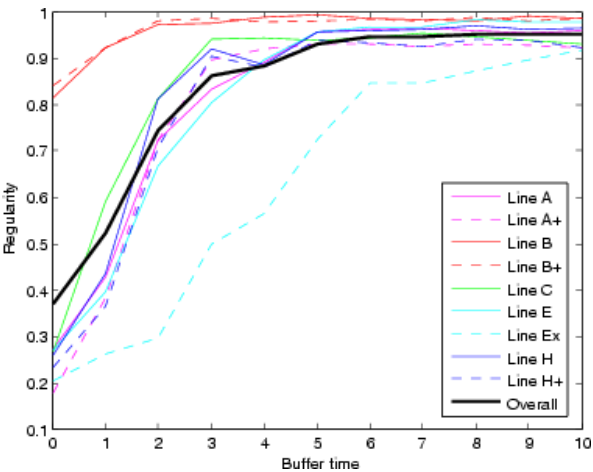


Figure C.11: Regularity on the lines of the timetable with 9 lines with different sizes of buffers on terminals

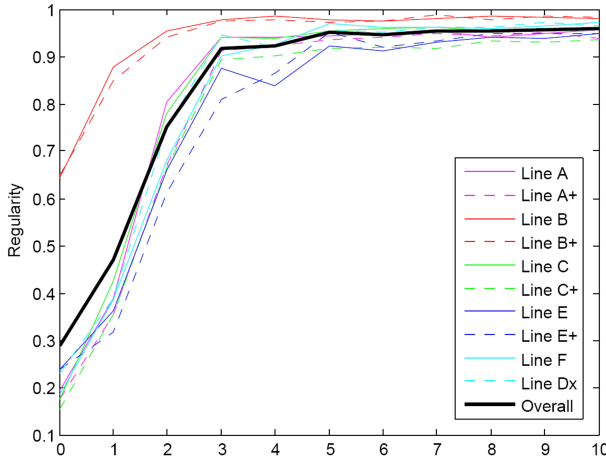


Figure C.12: Regularity on the lines of the timetable with 10 lines with different sizes of buffers on terminals

C.9 Extensions and concrete application

In fall 2006 DSB S-tog investigated a new concept for a timetable expected to be less sensitive to delays because of more homogeneous stopping patterns. The challenge for DSB S-tog in the process was to support this expectation of higher robustness and thereby improved customer satisfaction by quantitative evidence.

Traditional simulation tools used in the railway sector are very data-intensive and useful for detailed timetable design but not well suited for rapid comparison of different potential timetables. Hence with such detailed models, only a few scenarios and concepts are usually defined, tested (simulated) and compared. The simulation model described previously was therefore modified to allow for use in an analysis of the robustness of both the existing and the proposed timetables. The time necessary to set up, test, execute, evaluate and analyze different scenarios using the simulation model is significantly smaller compared to a highly detailed traditional model.

A project was established in corporation with PA Consulting Group, in which the model was extended to capture additional details. These were expected to have substantial impact on either the original or the new timetable concept. The extensions are:

- *Restrictions at the termini.* Crossing tracks (with the effect of single tracks) at the end stations are implemented. Hence, the time between trains leaving and entering the termini must be above a given thresholds. Furthermore, an upper limit is set for the capacity at each end station.
- *Recovery by holding back trains.* Trains stopping on all stations on the line are held back to let fast trains pass. (This is implemented at the stations København and Lyngby in both directions and stations Holte, Ballerup, Hundige for trains heading for the central area).
- *Recovery by gaining time at routes.* The trains can gain lost time on the routes between stations.
- *Optional choice of platform.* Change on platform according to schedule/sequence is made possible at København in both directions (the trains occupy the first free platform with no regard to the given sequence).
- *Single Track implemented.* The single track at Fiskebæk bro is implemented as an extra station. All trains on the lines to and from Farum visit this fictive station.
- *Variable headways.* Headways are variable according to the stations.
- *Delay distributions.* New delays distributions produced by DSB S-tog are used. The delay distributions are still station dependent.

The model was further more improved according to specific output needs. New key indicators for performance and diagnostics has been designed with the purpose of comparing and specifically capturing robustness, for example distribution of the trains in the network, time between departures, clustered departures, accumulated secondary delays and the use of slack. To improve the analysis phase the model is connected to an Access database in which all in- and output are stored and analyzed.

The simulation experiments conducted shows that the new concept can be expected to result in a much more robust timetable characterized by:

- *Higher regularity.* The regularity of the new timetable was much higher than that of the original timetable. This implied more precise and reliable departures.
- *Less accumulated secondary delays.* The accumulated secondary delay in the new timetable was only one third compared to the original timetable. This indicated fewer dependencies and higher stability.

- *Less use of added slack.* The utilization of the added slack was only half as high in the new timetable compared to the original timetable. This indicated that less slack time was needed to maintain stability and that fewer secondary delays were accumulated in the new timetable.

The positive effect of the homogeneous stopping pattern was confirmed. The original timetable has far more close departures with only a few minutes between them. Most of the secondary delays were initiated by trains with such close departures.

The model demonstrated a clear difference in robustness of the two timetables - and gave insight in the reasons through the diagnostic indicators. Hence, this tactical tool gives DSB S-tog the possibility of rapidly testing new ideas supporting the continuous development of new timetables concepts with the purpose of constantly optimizing robustness, maximizing rolling stock utilization (by minimizing slack and driving time) or minimizing waiting time for passengers.

C.10 Conclusions and future work

We have presented a simulation model for testing timetable robustness and the effect on robustness of three different recovery strategies. The main results from our tests are that there is a upper limit on the amount of buffer time leading to positive effect on the regularity, and that small delays though insignificant on their own have a significant additional effect when occurring together with large delays. Finally, there is a clear tendency that the recovery methods rendering the largest increase in headways result in the best robustness and thereby the best increase in regularity.

Further work on the simulation model is to implement various others of the presented recovery methods. Also, simulating the operation during non-peak hours including the implementation of rules for change of train-formation is of obvious interest. Furthermore, including the train drivers in the simulation will enable analysis of the dependency between timetables and crew plans, but will also require substantial additions and changes to the underlying model.

Bibliography

- M. Carey and S. Carville. Exact heuristic measures of schedule reliability. *Journal of Operational Research Society*, 51:666–682, 2000.
- M. Carey and S. Carville. Testing schedule performance and reliability for train stations. *International Transactions in Operational Research*, 11:382–394, 2004.
- C. J. Goodman and R. Takagi. Dynamic re-scheduling of trains after disruptions. *Computers in Railways*, IX, 2004. WIT Press.
- R. M. P. Goverde and M. A. Odijk. Performance evaluation of network timetables using peter. *Computers in Railways*, VIII, 2002. WIT Press.
- M. Hofman and L. F. Madsen. Robustness in train scheduling. Master’s thesis, The Technical University of Denmark, 2005.
- J. S. Hoogheimestra and M. J. G. Teunisse. The use of simulation in the planning of the dutch railway services. In *Proceedings of the 1998 Winter Simulation Conference*, 1998.
- T. Kawakami. Future framework for maglev train traffic control system utilizing autonomous decentralized architecture. *Autonomous Decentralized Systems*.
- W. D. Kelton, R. P. Sadowski, and D. T. Sturrock. *Simulation with Arena*. 3 edition, 2004.
- L. G. Mattson. Train service reliability: A survey of methods for deriving relationship for train delays. <https://users.du.se/jen/Seminarieuppsatser/Forsening-tag-Mattson.pdf>.

- D. Middelkoop and M. Bouwman. Simone: Large scale train network simulations. In *Proceedings of the 2001 Winter Simulation Conference*, pages 1042–1047. Institute of Electronical and Electronic Engineering, Piscataway, New Jersey, 2001.
- A. Puong and N. H. M. Wilson. A train holding model for urban rail transit systems. In *Proceedings of Conference on Computer Aided Scheduling on Public Transport*, 2004. <http://fugazi.engr.arizona.edu/caspt/puong.pdf>.
- Sandblad et al. T9 simulatorsystem inom tågtrafikstyring, en kundskabsdokumentation. Technical report, Upsala Universitet, 2003.
- J. C. M. Vroman, R. Dekker, and L. G. Kroon. Reliability and heterogeneity of railway services. Technical report, Erasmus University Rotterdam.

APPENDIX D

Optimal reinsertion of cancelled train line

IMM-Technical Report-2006-13.

Optimal Reinsertion of Cancelled Train Line

Julie Jespersen Groth^{a,b}, Jesper Larsen^b and Jens Clausen^{a,b}

^a DSB S-tog, Denmark

^b Informatics and Mathematical Modelling, Technical University of Denmark,
DK-2800 Kongens Lyngby, Denmark

Abstract

DSB S-tog (S-tog) is the operator of the suburban rail of Copenhagen, Denmark. The suburban network covers approximately 170 km of double-track and 80 stations. When larger disturbances occur in the S-tog network one of the countermeasures is to take out entire train lines. When the disturbance has been resolved the problem is to decide when to start the reinsertion at each rolling stock depot in order to resume scheduled service as early as possible. The process of resuming service is regulated by a number of constraints. Hence, the task of calculating a reinsertion plan of a train line is complex. Here we present a mixed integer programming model for finding a reinsertion plan of a train line minimizing the latest time to reinsertion. The MIP model has been implemented in GAMS and solved with Cplex. The optimal solution is found within an average of 0.5 CPU seconds in each test case. Reinsertion plans in operation is today determined by the reinsertion model.

D.1 Introduction to DSB S-tog

DSB S-tog (S-tog) is the operator of the suburban rail of Copenhagen, Denmark. The suburban network covers approximately 170 km of double-track and 80 stations. S-tog daily transports approximately 30.000 passengers. The tracks are controlled by the infrastructural owner BaneDanmark (BD), and S-tog is the only user.

The S-tog network consists of train lines covering the S-tog infrastructure by various routes depending on the timetable in use. Figure D.1 shows the present lines in the network. The network can be thought of as consisting of 8 sections;

A central section, 6 "fingers", and a circular rail (lines F and F+ in yellow). Except for the circular rail the lines merge in the central section, and they split as they enter the fingers according to their schedule.

The structure of the S-tog network implies that 10 lines intersect in the central section. The trains on each train line run with a frequency of 3 trains per hour resulting in an frequency period of 20 minutes. Hence, in every 20 minutes interval there is on average only 2 minutes between each train in the central section i.e. there is a 2 minutes average headway between the trains. Such low headway implies that even small delays often have a significant negative effect on a substantial number of trains.



Figure D.1: The network of DSB S-tog in 2006

Each line in the S-tog network is covered by 4 to 10 trains depending on the duration of the line circuit, which is the time it takes for a train to drive from one terminal to the other terminal and back. The duration of a circuit and the time planned for turnaround at the terminals divided by the frequency period (20 minutes) determines the number of trains required to run the line. Each train consists of one or more train units. The composition of a train varies during the day according to the expected passenger demand.

At all times a train number is associated with each operating train. The train

number is changed every time a train turns at a line terminal to run in the opposite direction. For each train there is hence a series of train numbers during the day defining the train tasks of that particular train. A train task is defined by a departure from a terminal s_1 at time t_1 and the subsequent arrival at a terminal s_2 at time t_2 . The number series of a train is called the *train sequence* and there can be only one train sequence for each train. Also, two train numbers cannot occur in two different train sequences. Figure D.2 shows an example of a line covered by two trains where each train is covered by a train sequence and different train units during the day.

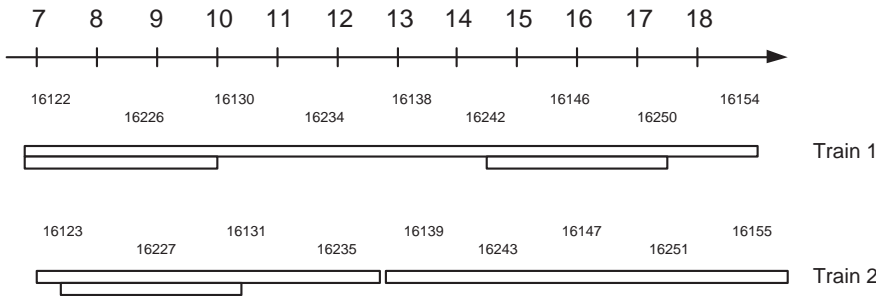


Figure D.2: Two train series together covering the two trains on a train line (here represented by the train numbers). Each train number indicates a train task for the train. The blocks illustrates train units covering the task. When two blocks cover the task there are two train units assigned to the task. Hence, two units are needed in the morning rush hour for both trains. Only train 1 is covered by two units in the afternoon.

There is much information embedded in the train numbers. Each train number identifies the present train line, stopping pattern, direction (north/south) and an time interval of arrival at KH for a train. The time interval is given by the two last digits in the train number e.g. if these are 35 the integer part of a division by 3 indicates that the train arrives at hour 11. The remainder, which is 2, indicates that the train arrives at the 3rd frequency period within that hour. For two train numbers on the same train line, the train number with the lowest value of its two last digits will be the train task performed first. For two adjacent train numbers in the same train sequence there is a constant interval between the values of the last two digits.

During the daily operation incidents occur that disturb the scheduled departures. One way to compensate for potential disturbance is to construct timetable, crew, and rolling stock plans with included buffer times. For example, in the timetable buffer times can be included in the headways, the dwell times on sta-

tions and as a part of the turnaround times. It is not necessarily evident where in the schedules it is optimal to allocate the buffers. This information can be derived by e.g. simulation studies based on real observations of the schedules, c.f. Hofman et al. Hofman et al. [2008] and e.g. Kroon et al. Kroon et al. [2005].

Even though precautions are taken to minimize the effects of incidents, it is not possible to avoid delays completely.

When larger disturbances occur on the S-tog network, the disturbance is often redeemed by taking out an entire train line i.e. all departures on a train line are cancelled. By taking out a train line more slack is created in the timetable, i.e. the headways are increased between time adjacent train lines. This creates increased buffer times in the timetable and more room is created for absorbing the delays.

A cancellation of a train line is implemented by shunting the rolling stock to depot tracks as the trains arrive at rolling stock depots. Rolling stock depots are located at most of the line terminals and at the central station. In the process of cancellation it is normally not allowed to drive “backwards” in the network i.e. driving a train in the opposite direction of what it was originally planned to drive. A train can only move forward to the next depot where it will be taken out of service. Therefore, the trains on the cancelled train line ends up being distributed among the depots along the line according to where they were in the network when the decision of cancelling the line was made. Recall that a train is defined by its train sequence and not by the train units covering it. Hence, the train units that are taken out at a depot are not necessarily used to cover the same trains when reinserted.

When an adequate level of punctuality has been re-established in the operation, the cancelled train lines must be reinserted according to the original timetable. The reinsertion is governed by a set of rules and has to be performed as early as possible while respecting these. This paper addresses the problem of determining the reinsertion scheme such that the latest reinserted train is inserted as early as possible.

The remainder of the paper is organized as follows. In Section D.2 we describe the reinsertion problem in more detail. Section D.3 presents a mixed integer programming model for the problem. Computational results are presented in Section D.4. Section D.5 describes the initial background for formulating a mathematical model for the problem.

Finally potential further developments and a conclusion are given in Section D.6 and D.7.

D.2 The Reinsertion Problem

The status of the running operation is evaluated by a train leader from the infrastructure owner, BD. The train leader has the final responsibility of operation. After the decision of initiating a reinsertion has been made, the reinsertion must be carried out as quickly as possible.

When a parked train is reinserted it is transported as empty stock from the depot tracks to a platform by shunting personnel. A train driver for the train arrives on an operating train from the crew depot at the central station (KH). The train to be reinserted must depart according to a scheduled departure on the relevant train line.

For the reinsertion problem it is assumed that train units will be shunted to departing platforms according to standard shunting times and available track routes. The problem is then to assure that all trains are inserted at some depot on the train's route.

The reinsertion plan is calculated by a rolling stock dispatcher from S-tog. The reinsertion is presently computed for one train line at a time. It is necessary to decide which trains already in operation can transport train drivers to the rolling stock depots, from where the trains are inserted. The number of trains to be inserted from each depot is determined by the dispatcher, however, it is not given which train units at the rolling stock depots should be inserted to cover which trains in the schedule of the train line.

For most lines, intermediate rolling stock depots exist along the route of the line. As for the terminal depots, it is determined, prior to the calculation of a reinsertion plan, how many trains must be inserted from the intermediate depots in total. At an intermediate depot trains can be inserted in both directions of the line. Inserting in both directions decreases the finishing time of the reinsertion process. Two departures in opposite direction are not bound by headways as there is generally not interdependency between the infrastructure going north respectively south. Therefore, the total time of insertion can be reduced by inserting in both directions on an intermediate depot. As an example, consider three trains to be inserted from a depot. Departures going north are at minutes 03, 23 and, 43 every hour. Departures going south are at minutes 06, 26 and, 46 every hour. If all trains are inserted in one direction, the reinsertion will span 40 minutes. If the trains are inserted in both directions, the reinsertion will span just 20 minutes.

In the reinsertion problem each train must be reinserted only once. Also, the reinsertion must be made under two different considerations of order. Firstly, if

reinsertion has begun from a given rolling stock depot, the remaining trains to be inserted from that depot must be inserted in order according to the frequency period. For example, at S-tog the frequency period is 20 minutes on all train lines. If 3 trains must be reinserted from Farum rolling stock depot and the first reinserted train departs at 15:18, then the remaining 2 trains must be reinserted and depart at respectively 15:38 and 15:58. Inserting the remaining two trains at 15:58 and 16:18 leads to an unassigned departure at 15:38, which is an illegal solution. Secondly, the order with respect to frequency must also be kept across rolling stock depots. That is, after the initiation of reinsertion, the time between two adjacent departures on any station in the network must always be the frequency period of 20 minutes.

One of the advantages of the reinsertion model is the solution time of the model compared to manual calculations. Also, it is possible to calculate a reinsertion plan immediately when the distribution of trains among depots is known after the take out and a decision regarding the number of trains to reinsert at each depot has been taken. As the timetable is periodic the reinsertion plan calculated will in principle be identical except for the exact train numbers to be inserted. This may lead to some advantages with respect to coordinating the train driver schedules according to the reinsertion, thereby preventing reinsertion schedules being discarded because of the lack of drivers.

D.2.1 References

Comment in more detail on the contents of these. Recent surveys on rail operation models are given by Cordeau et al. [1998], Huisman et al. [2005] and Törnquist [2006]. The first survey considers railway operation specifically from the freight operator's side. In the second operation research problems of passenger railway operation in especially Europe is described in all planning phases from strategic to short term. Finally, the third survey the railway dispatching models from the infrastructure perspective considering the dimensions of the railway network e.g. number of parallel tracks considered in the models.

The reinsertion problem and models for solving it is not mentioned in either of these surveys. A thorough search has not produced any additional literature that resembles the problem of reinserting train lines.

D.2.2 A real life example

To illustrate the reinsertion problem we give an example of a reinsertion. Two lines, H and H+, run on the route between Frederikssund (FS) and Farum (FM). When large disturbances occur involving the sections of this route, the H+ line is typically taken out. Each of the 10 trains servicing the H+ train line are taken out on one of the terminal rolling stock depots, FS and FM, or on the intermediate depots at Ballerup (BA) and KH. In this example, 2 trains are taken out on each of the terminal depots and 3 on each of the intermediate depots. One scenario of reinsertion is then that with respect to trains, reinsertion is carried out exactly according to where these have been taken out, in which case 2 trains must be reinserted from each terminal depot and 3 trains must be reinserted from each intermediate depot, where insertion is possible in both directions. If extra train units are available on any of the depots several reinsertion scenarios are possible according to how many units are available and at which depots.

Figure D.3 illustrates the trains that are available for transporting drivers from the crew depot at KH to the respective rolling stock depots. Train *a* and *d* are the first such trains going respectively south and north from KH that can bring out drivers to depots. As these trains pass the depots in each direction potential drivers can be made available for reinsertion.

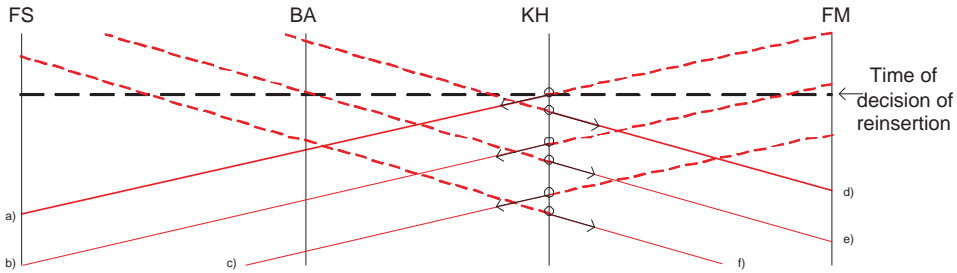


Figure D.3: The straight lines a, b and c illustrates trains carrying drivers going south and the lines d, e and f those going north. The lines initiate reinsertion at the different depots as they pass them.

In the reinsertion model the initiation time of reinsertion on each depot is counted in integral time slots according to the frequency period within the timetable. It is counted how many trains on the train line in question was planned to leave the depot from the decision of reinsertion until the first driver-carrying train reaches the depot. In Figure D.5 there are 2 trains originally planned to leave the FS depot before reinsertion can begin. Notice that the

time of decision of reinsertion is coinciding with the departure time of the first south going train being able to carry drivers. Formally, this is always the case.

The exact approach of a reinsertion is illustrated in Figure D.4. For each of the figures a) - d) trains are inserted from a depot. Observe that order is kept at all time. There is no vacant frequency periods at depots and there are no stations where passengers experience vacant frequency periods. Illustrated in red on a), b) and d) is the driver-carrying trains transporting drivers for the reinsertion.

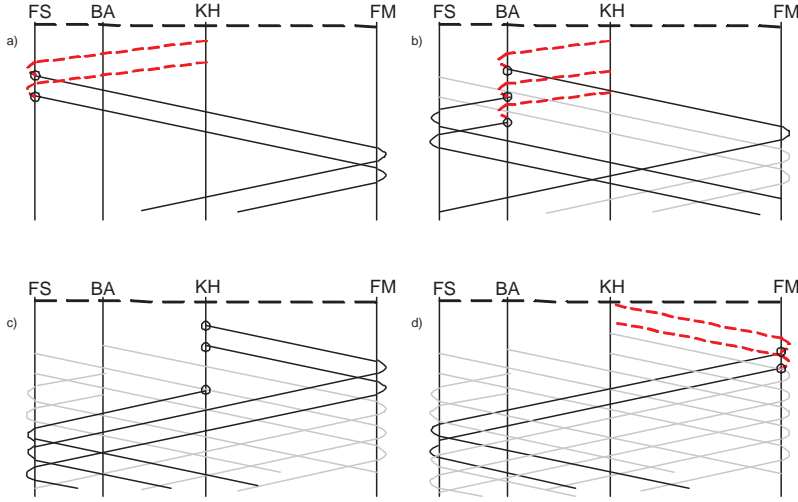


Figure D.4: Illustration of a reinsertion. In a) trains are inserted from depot FS, in b) from depot BA, in c) from KH and in d) from FM.

D.2.3 Input to the reinsertion model

The solution to the reinsertion problem is based on a relatively small amount of information. There are two types of input necessary in the model: input based on background data and input based on real time data. Firstly, the model must be built with background data based on the timetable structure. Secondly, after the construction of the model certain input in real time is necessary for deriving the right reinsertion plan according to the relevant real time scenario.

When building the model it is essential to know for each depot how many cancelled departures there are from the time of the decision to start reinsertion

and until the first driver is ready for driving the first potential reinserted train, see Figure D.5. The time of decision of reinsertion is always coinciding with the departure of a train from KH, which can carry drivers to the reinsertion depots. Furthermore, it is necessary to know the number of trains on the train line. As mentioned in Section D.1, each train can be viewed as a sequence of train numbers. Due to the periodic format of the timetable the solution to the reinsertion problem is generic i.e. the structure is independent of the specific times given in the timetable. Therefore, when building the model, it is only necessary to be able to differ between the trains. It is sufficient to make one calculation for each distribution of trains over depots. As there is only a limited number of possible distributions of trains among depots, all solutions can in fact easily be generated in advance and updated according to time of day in the real time situation. Knowledge of exact train numbers to be reinserted are not relevant before real time.

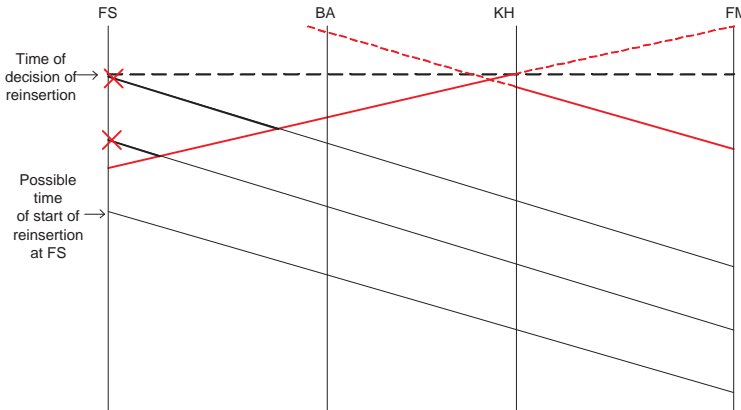


Figure D.5: Illustration of the reinsertion start time at the FS depot. From the decision of reinsertion until the reinsertion can begin at the FS depot, the two first trains with scheduled departures after the time of decision of reinsertion cannot be reinserted as the train carrying drivers has not yet reached the depot.

In real time the necessary input is the train distribution among the depots and which operating train number is the first that can carry drivers to the depots. The specific train numbers to be inserted can also be derived from the train number given as input and the reinsertion solution, which can be looked up based on the distribution of trains on depots. Additionally, the train numbers that will be used to transport the train drivers to the rolling stock depot must be identified. These can also be derived from the reinsertion solution and the input train number. Summing up, the solution looked up by the rolling stock

dispatcher is used to find the train numbers of respectively the trains to be reinserted and the trains to transport drivers.

D.3 The Reinsertion Model

Let I be the set of trains that must be inserted and K the set of depots they can be inserted from. J is the set of available time slots for reinsertion. The goal of the model is to decide which train, $i \in I$, should be inserted from which depot, $k \in K$. Each originally scheduled train i (before cancellation) must be covered with train units and hence reinserted in operation according to schedule. Also the model must decide for each train in which time slot, $j \in J$, the reinsertion will take place.

The model decides which trains will run but it does not consider which specific train units to use to cover the trains. It is assumed that the information on the distribution of train units across depots, $D_k, k \in K$ is provided as input and thereby the train units are sufficient in number to cover the trains.

The variables representing which train to be inserted from which depot and when are binary. Trains are indexed by i , time slots by j , and depots by k .

$$x_{ijk} = \begin{cases} 1 & \text{if train } i \text{ is inserted in time slot } j \text{ from depot } k \\ 0 & \text{otherwise} \end{cases}$$

To ensure that each train is inserted in an correct time slot, it is necessary to take into consideration the train sequences of each train describing in which time slot each train is at the different depots. To handle this a constant is introduced, in_{ijk} .

$$in_{ijk} = \begin{cases} 1 & \text{if train } i \text{ may depart from depot } k \text{ in time slot } j \\ 0 & \text{otherwise} \end{cases}$$

It is not possible to insert a train from a depot, if it is not there at that specific time slot. We refer to this as the order between stations and it is ensured by inequality (D.1).

$$x_{ijk} \leq in_{ijk}, \quad \forall i \in I, j \in J, k \in K \quad (\text{D.1})$$

Each train must be covered exactly once. This is guaranteed by the partitioning constraints (D.2)

$$\sum_{j,k} x_{ijk} = 1, \quad \forall i \in I \quad (\text{D.2})$$

Inequalities (D.3) are included so that no time slot for a depot or train can be covered more than once:

$$\sum_i x_{ijk} \leq 1, \quad \forall j \in J, k \in K \quad (\text{D.3})$$

The number of trains to be inserted from each depot is known. Therefore, binding constraints exist for each depot. They differ for respectively terminal and intermediate depots. As the trains are inserted only in one direction at the terminal depots, $k \in K_T$, the binding constraints for these depots are :

$$\sum_{i,j} x_{ijk} = D_k, \quad \forall k \in K^T \quad (\text{D.4})$$

As mentioned earlier a better solution can be achieved when insertion at intermediate depots are made in both directions. According to current practise half of the trains on each intermediate depot are inserted in one direction, the other half in the opposite direction. This is handled in the model by including two depots for each intermediate depot. The set of intermediate depots is denoted K^I . It is constructed by sets of two depots together denoting one intermediate depot where reinsertion can be carried out in l directions, $K^I = K_1^I \cup \dots \cup K_l^I$, where $l \in L$. L is the set of directions, which in the S-tog network for all depots is north or south. The total set of depots is $K = K^T \cup K^I$. Decision variables $D_k^I, k \in K^I$ have been added to the model to determine the number of trains inserted each direction.

The sum of trains inserted in both directions should equal the total number of trains to be inserted from the intermediate depot. Equations (D.5) ensure that the number of trains inserted in each direction is the total number of trains to be inserted divided by 2. If an odd number of trains is to be inserted, the result is rounded up or down to nearest integer depending on which is more favorable to the model.

$$\sum_{i,j,k} x_{ijk} = \sum_k D_k, \quad \forall l \in L, k \in K_l^I \quad (\text{D.5})$$

$$\sum_{i,j} x_{ijk} = D_k^I, \quad \forall \quad k \in K^I \quad (\text{D.6})$$

$$D_k^I \geq \lfloor \frac{D_k}{2} \rfloor, \quad \forall \quad k \in K^I \quad (\text{D.7})$$

$$D_k^I \leq \lceil \frac{D_k}{2} \rceil, \quad \forall \quad k \in K^I \quad (\text{D.8})$$

As mentioned in Section D.2 order must be kept within depots and between depots. Also, reinsertion can not begin on a depot before a train driver has arrived from the crew depot to drive the train to be reinserted.

Recall that no vacant frequency periods can occur in a feasible solution. To model this, we introduce two sets of integer variables, $start_k$ and end_k . Also, we introduce equations (D.9) to (D.12). Equations (D.9) connect the start and end variables. Equations (D.10) assure that reinsertion is not begun before the first driver can arrive at the depot. The constant, C_k , indicates how many trains have been scheduled at depot k from the time of the decision of reinsertion until drivers are able to reach the depot, cf. Figure D.5. Equations (D.11) and (D.12) ensure that when a reinsertion has begun on depot, it is carried out continuously in adjacent time slots.

$$start_k + \sum_{i,j} x_{ijk} - 1 = end_k, \quad \forall \quad k \in K \quad (\text{D.9})$$

$$start_k \geq C_k + 1, \quad \forall \quad k \in K \quad (\text{D.10})$$

$$start_k \leq j + M \cdot (1 - x_{ijk}), \quad \forall \quad i \in I, j \in J, k \in K \quad (\text{D.11})$$

$$end_k \geq j - M \cdot (1 - x_{ijk}), \quad \forall \quad i \in I, j \in J, k \in K \quad (\text{D.12})$$

At S-tog much of the information of the timetable and departures is embedded in the train numbers. The periodic form of the timetable supports this formulation. The train numbers to be inserted when x_{ijk} is 1 is calculated from an initial train number on a train able to carry train drivers to the depots and some

constant describing the relationship between the train numbers on the driver-carrying line and the line to be reinserted. The train number, tn_{ijk} , is adjusted according to the time slot in which it is to be inserted. See equation (D.13).

$$tn_{ijk} = \left(InitialTrain + TrainConst_k + j \right) \cdot x_{ijk}, \quad (D.13)$$

$$\forall i \in I, j \in J, k \in K$$

The objective of the model is to reinsert so that the last reinsertion is performed as early as possible. This is assured by an objective function of minimizing the maximum inserted train number, Z . As information of time of day is embedded in the train numbers this equivalent to minimizing the latest time of reinsertion. This is achieved by minimizing the maximum value of the last two-digit number in the train number.

$$Minimize \quad Z \quad (D.14)$$

$$Z \geq tn_{ijk}, \quad \forall i \in I, j \in J, k \in K \quad (D.15)$$

D.4 Computational Experience

Test results show that the running time of the model on average is only approximately 0.5 CPU seconds, i.e. the model solves the problem in real time for the relevant problem instances. The real time approach is, however, not chosen partly due to software license issues, partly due to the generic nature of the reinsertion plans. These are hence generated in advance and looked up by the rolling stock dispatchers at the relevant point in time.

If, for example, $K = \{FS, BA, KH, FM\}$ and $D_k = \{2, 3, 3, 2\}$ the optimal reinsertion scheme is illustrated in Table D.1.

The practically applicable solution can be derived based on the train number, 50227, the reinsertion solution above and a set of constants. Suppose that we wish to reinsert a set of cancelled trains starting at 9 am. The first train able to transport drivers south going leaves at 9:08 and has the train number 50227. We assume that the distribution of trains on depots is as indicated above. The practically applicable solution corresponding to Table D.1 is given in Table D.2.

Station	Train	Inserted in time slot:
FS	1	4
KH south going	2	2
KH south going	3	3
BA south going	4	5
FM	5	3
FM	6	4
KH north going	7	3
BA north going	8	2
BA north going	9	3
FS	10	3

Table D.1: A solution derived by the model. The first column shows the depot and direction of insertion. The second column shows which of the trains are inserted. The last column shows in which time slot at the depot the train is inserted.

Station	Train	Train number to transport drivers:	Train Number to insert
FS	1	50228	55133
KH south going	2	Driver present	55228
KH south going	3	Driver present	55229
BA south going	4	50230	55230
FM	5	50127	55231
FM	6	50128	55232
KH north going	7	Driver present	55129
BA north going	8	50227	55130
BA north going	9	50228	55131
FS	10	50227	55132

Table D.2: A practical applicable solution. The first column shows the depot and direction of insertion. The second column shows the train number of any driver carrying train. The last column shows the train number to be inserted.

Equation D.13 is used to calculate the train numbers to insert. For example, for the first row in Table D.1 $i = 1$, $j = 4$ and $TrainConst_{FS} = 4902$ which in the first row and last column of Table D.2 gives the train number 55133. $TrainConst_{FS}$ is adjusted to give the right direction, line and time of the inserted train number. The train number carrying drivers to the insertion depot is calculated by an equation similar to (D.13) with other constants.

Each train number indicates a time and a direction. This is sufficient knowledge for the dispatchers to carry out the reinsertion.

D.5 An Improved Planning Process

The initial request for a tool for calculating reinsertion was made by the rolling stock dispatchers at S-tog. They found the problem of creating reinsertion schedules by hand as complicated and time demanding, especially in real time where time is sparse. First a tool was made that was not based on the principles of MIP. It was merely a spreadsheet calculating the reinsertion plan from basic knowledge of the distribution of trains, the first driver-carrying train and a large set of if-then-else-loops. The project of creating an optimization model for calculating the reinsertion was started mainly due to the quite complicated task of updating the initial reinsertion tool. The mathematical model of the reinsertion problem has been implemented in GAMS and solved in CPLEX. It has been in operation since the August 2006.

Solutions are generated with the MIP model for all possible scenarios of train distributions over rolling stock depots. The solutions are then stored and the rolling stock dispatcher can look up the solutions via a spreadsheet.

The reinsertion model has increased the level of service offered to the passengers. Earlier, when the rolling stock dispatcher had to make the calculation of reinsertion by hand, the solutions were either not generated because it took too long time to calculate a solution, or a solution was generated with a longer total reinsertion time than the optimal solution. In the first case the train lines would remain cancelled for the remainder of the day.

Besides the passenger service improvement, the reinsertion model decreases the level of stress for the rolling stock dispatchers. Solutions can be generated immediately to satisfy the demand of the train controllers in charge of the reinsertion decision. This has resulted in a more efficient planning process with resources left for other tasks.

Furthermore, the MIP model is easy to update according to a new periodic timetable. This is done simply by changing the set of constants presently used in the model and generating a new set of solutions.

D.6 Further Developments

Presently the reinsertion model is used only for scenarios where a cancelled train line needs to be inserted into a running operation, in which running trains can transport drivers to rolling stock depots. Future developments on the model may include complete start-ups where trains can be inserted as the first on their route.

The present model works with a distribution of trains reinserted in each direction on intermediate depots of half reinserted in one direction and the other half reinserted in the other direction. Further developments on the model involves changing the constraints (D.7) and (D.8) to enable solutions where the number of trains reinserted in each direction on each intermediate depot are not bounded to be the half of the total number of trains to be inserted.

The number of trains to be reinserted on each depot is input to the model. Occasionally the number of trains available for operation on the depots is larger than the number of trains that

has been taken out. It seems natural to change the model in order to account for this fact by deciding the optimal number of trains to be inserted from each depot, ensuring that the total number of trains reinserted is the number of trains needed to cover the line.

At some of the routes of the S-tog network more than two lines cover a route simultaneously. A relevant recovery scenario is that more than one line is cancelled along the route. It is an obvious idea to modify the reinsertion model so that it can solve the problem of reinserting more than one line at a time. In the model this can be achieved mainly through a modification of input.

D.7 Conclusion

We have presented a MIP model for generating optimal reinsertion plans of an entire, cancelled train line which is employed in operation today. Optimality is

often not achieved when the reinsertion is calculated by hand by the dispatchers in an often stressed situation. As a consequence, train lines sometimes in the past remained cancelled for the remainder of the day even though this is not necessary.

The solutions generated by the reinsertion model are fully applicable in operation i.e. it is non-complex to derive a practical solution from a feasible solution found by the model. The MIP model is easy to update according to a new periodic timetable.

Earlier, the difficulties in calculating a reinsertion plan prevented different factors from being taken into consideration cf. Section D.5. The significantly decreased solution time of the reinsertion problem (when comparing solutions calculated by hand and solutions generated with the reinsertion model) gives the possibility of adjusting the reinsertion carried out in operation.

Bibliography

- J.-F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32:p. 380–404, 1998.
- M. Hofman, L. F. Madsen, J. J. Groth, J. Clausen, and J. Larsen. Robustness and recovery in train scheduling – a simulation study from dsb s-tog a/s. *International Journal of Operations Research*, 2008. Accepted for publication.
- D. Huisman, L. G. Kroon, R. M. Lentink, and M. J. C. M. Vromans. Operations Research in passenger railway transportation. *Statistica Neerlandica*, 59(4): p. 467–497, 2005.
- L. G. Kroon, R. Dekker, and M. J. C. M. Vromans. Cyclic Railway Timetabling: a Stochastic Optimization Approach. In F. Geraets, L. G. Kroon, D. Wagner, and C. Zaroliagis, editors, *accepted for Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science*, Berlin, 2005. Springer.
- J. Törnquist. Computer-based decision support for railway traffic scheduling and dispatching: a review of models and algorithms. In *Presented at ATMOS2005, Palma de Mallorca, Spain, October 2005 (proceedings to be published on the Dagstuhl Research Online Publication Server (DROPS) during 2006*, 2006.

APPENDIX E

The rolling stock recovery problem

IMM-Technical Report-2008-18.

Rolling Stock Recovery Problem

Julie Jespersen Groth^{a,b}, Jens Clausen^{a,b} and Jesper Larsen^b

^a DSB S-tog, Denmark

^b Informatics and Mathematical Modelling, Technical University of Denmark,
DK-2800 Kongens Lyngby, Denmark

Abstract

Real time decision support in railway operations is an area which has so far received limited attention. In this paper we address real time recovery of a rolling stock plan. Given a disturbed rolling stock plan the objective is to return quickly and inexpensively to the original rolling stock plan. Each train unit is hence rerouted through the train network so that each terminal departure is covered sufficiently wrt. seats relative to demand and so that the train unit paths are feasible with respect to connections.

We address the rolling stock recovery problem using a method based on decomposition where first the number and order of train units for each departure are determined. Given this knowledge we find the train path for each train unit. The experimental results show promising solution times and quality indicating applicability in practice.

E.1 Introduction

During the last years there has been an increased focus on developing tools to aid the planning process in railway transportation. The tools are computer software, which can fully or partially automate parts of the planning process. As in other industries the initial focus has been on strategic, tactical and operational planning. Only lately focus has turned to the area of short term and real time planning. This paper concentrates on the area of rolling stock real time planning. All models are based on the suburban railway network in Copenhagen, Denmark. The railway operator operating the network is DSB S-tog A/S.

The areas of operational, short-term and real time planning, can with respect to rolling stock be described as follow.

Operational The operational planning process is based on the tactical plan, which defines the number of train units and which type is assigned to each defined train task. A train task in this context is defined by a departure from a station and an arrival at another station. The stations most often are rolling stock depots. Rolling stock unit types are assigned to train tasks in such a way that later, train unit routes can be build for physical units that enables each train unit to visit the maintenance center within the predefined safety time and kilometer limit.

Also, in operational planning adjustments are made with respect to infrastructure maintenance works.

Short-term Short-term planning in the railway business concerns the routing of the physical train units approx. 1-3 days in advance of operation. Also in this phase small adjustments to the number and type of train units assigned to each train task may be necessary.

Real time The major difference from operational to short-term planning of rolling stock is that for the latter information of the physical train ID's are included. This level of detail is maintained also in real time. Real time planning is conducted during the operation. Real time rolling stock planning is the re-planning or recovery of the plan for physical train units after disruption has occurred. This is also called rolling stock disruption management.

In practice rolling stock dispatchers monitor the operation of the rolling stock plan and the depot plans. In the cases where the operation does not run according to the rolling stock plan, the rolling stock dispatcher makes real time decisions on the re-assignments of train units to train tasks. Often suboptimal decisions are made due to the complexity of the task of manually establishing an integrated solution taking into consideration the recovery of several trains.

There is a substantial cost of re-allocating train units after a disruption in the rolling stock plan. The reallocation is necessary to meet end-of-day depot balance requirements and the maintenance requirements of each individual train unit. Furthermore, if too many train units are allocated to trains ending up at particular depot there may not be sufficient physical space in the depot to park all the train units.

Today, when a disruption has occurred the depot balances are often off implying that the rolling stock plan for the following day is also disrupted. Thus, either some train task must be covered insufficiently or not covered at all resulting in a cancellation of the task.

Next, in Section E.2 a review of related literature is given. In Section E.3,

we give an introduction to the terms of rolling stock planning. Hereafter, in Section E.4 we define terms concerning disruption. In Section E.5 the network is described. We introduce the Train position model in Section E.6, in Section E.7 the Train sequence model is presented and in Section E.8 the Train unit routing model is presented. Finally, in Sections E.9 and E.10 we present the Computational results and give a conclusion.

E.2 Literature review

The research within the area of rolling stock schedule optimization has up to recently mainly focused on the planning phases prior to the day of operation. Only little emphasis has been on the area of real time rolling stock recovery, see Nielsen [2008]. Huisman et al. [2005] give a survey on state-of-the-art Operations Research methods for solving passenger railway related planning problems. The real time handling of rolling stock is briefly mentioned and reference is made to the problems of short time planning, which resembles the real-time situation. Short-term rolling stock planning is done on a day-to-day basis, also adjusting the rolling stock plans according to changes in the timetable due to e.g. rail network maintenance work, or adjusting according to passenger flows, which may have changed the need for rolling stock assigned to each train task.

Other recent surveys on rail operation models are given by Cordeau et al. [1998], and Törnquist [2006].

At S-tog, the depots are physically not very large, and only one workshop is available for maintenance checks. Already in the initial operational rolling stock plan, the paths for the train units lead them pass to the workshop at regular intervals in time and distance.

The problem of planning rolling stock can be divided into two subproblems: Firstly, finding the compositions for each train task in the network and secondly, finding the paths for each virtual train unit ensuring depot feasibility and regular maintenance checks. The compositions indicate the type, number and order of train units assigned to a train task. The paths ensure that all train units are routed to pass the workshop at regular intervals.

The first problem of determining compositions is widely explored. There is a distinction between the problems of allocating rolling stock when the fleet is composed by train units compared to when it is composed by train carriages and train locomotives. Papers concerning the locomotive scheduling problem are Cordeau et al. [2001], Lingaya et al. [2002] and Brucker et al. [2004].

The first paper concerning the problem with self-propelled train units is Schrijver [1993]. In this paper a minimum circulation of rolling stock on a single train line running from Amsterdam to Vlissingen and vice versa is determined. The objective is to ensure sufficient seats available for each train task. The model does not take the train unit order within a composition into account. The problem is solved with commercial software for respectively one and two train unit types.

In Ben-Khedher et al. [1998] the problem of capacity adjustment is discussed. It is based on the problem of finding railway capacity for high speed trains running in the TGV network of SNCF, France. The model is based on the seat reservation system and the objective is to maximize expected profit.

Alfieri et al. [2006] address the problem of constructing circulations of train units. Focus is again on a single line. The model couples and decouples train units from trains as the depots are passed. The order within each composition is taken into consideration. The model is tested for two train types. The solution approach is based on a hierarchical decomposition into sub problems. First, the model, not taking compositions into consideration, is solved. Second, it is checked whether there is a feasible solution for the composition problem.

Peeters and Kroon [2003] present a branch-and-price algorithm for solving the allocation of train units to a single line or a set of interacting train lines. The model is tested on several real-life instances of the railway operator, NS Reizigers. Objectives considered are those of minimizing train unit km shortage, minimizing number shunting operations and number of driven train unit km. The model is based on a transition graph as is the model described in Alfieri et al. [2006]. The authors apply a Dantzig-Wolfe decomposition, reformulating so that a variable is associated with each path through the transition graph of all trains.

In Fioole et al. [2006] a model for finding the compositions of train units on train tasks is presented. Each solution is feasible with respect to composition order in depots and with respect to depot capacities. The model additionally takes into consideration combining and splitting of trains in depot junctions. It is an extension of the model described in Peeters and Kroon [2003]. The objective considers minimizing with respect to efficiency, service and robustness. The model is implemented and solved in the commercial integer programming solver CPLEX. This procedure improved the solution used in practice with up to 6 % with respect to number of driven train unit kilometres.

Given that the composition problem is solved at short term or real time level the problem of finding paths resembles the problem of finding work plans (lines of work) for crew. The train tasks form a time and space restricted path. Extensive

research within the area of crew planning has been carried out. Within the area of rail we refer to the survey of Huisman et al. [2005].

In Nielsen [2008] a generic framework for modelling the real time rolling stock re-scheduling problem is described. This is the problem of re-balancing the use of rolling stock on train tasks in real time. Rolling stock is considered at train type level. The modelling is based on the composition model presented in Fioole et al. [2006] and expanded to consider the end-of-day balances of rolling stock. The model have the objectives of minimizing number of cancelled trips, changes to the rolling stock depot plans and the end-of-day off balances. The model is solved using CPLEX 10.1. The registered computation times varies from few seconds up to a minute depending on the problem instances solved. All computational results are based on data from the Dutch railway operator NS Reizigers.

A recent paper, Rezanova and Ryan [2006], on the Train Driver Recovery Problem approaches the problem of recovering a train driver plan in real time given that some disturbances have disrupted the plan. The problem is solved using a set partitioning formulation. Fractional solutions for the LP relaxation of the IP problem is solved used constraint branching, however, most solutions are integer due to strong integer properties of the model. Solutions are found within few seconds.

Another interesting paper on railway recovery is Walker et al. [2004]. In this paper a model is described for simultaneous recovery of the train timetable and the corresponding crew plan. Promising results are presented for a single line of a New Zealand operator.

The current paper addresses the area of real time rolling stock recovery. No prior research is available on this subject. We introduce a decomposition method for the problem which provides good quality solutions quickly.

E.3 Basic elements of a rolling stock plan

Train operation runs according to a timetable consisting of terminal departures with predefined stopping patterns. Terminal departures are assembled in *Trains*. Each train is represented by a set of *Train tasks* forming a *Train sequence*, see figures E.1(a) and E.1(b). The train tasks of a train sequence form a predefined work plan for the train in which each train task, except for the first and the last, have a known predecessor and successor. This means that for two subsequent tasks t_1 and t_2 , $ArrivalTime(t_1) < DepartureTime(t_2)$ and

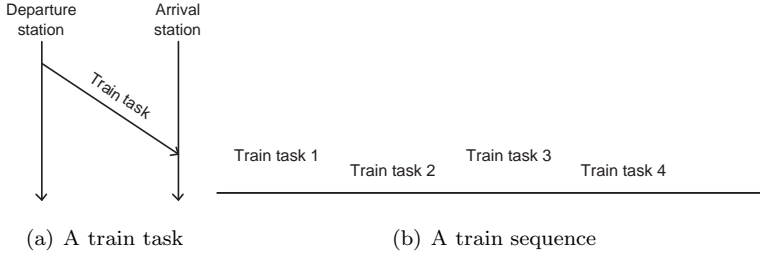


Figure E.1: Illustrating train terms

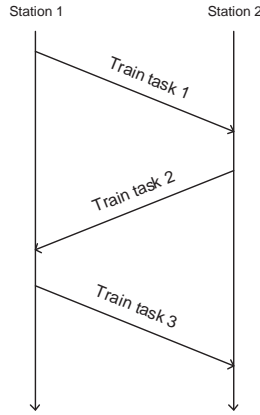


Figure E.2: Illustration of a train sequence in a time-space network

$ArrivalDepot(t_1) = DepartureDepot(t_2)$, see figure E.2. In the models presented later in this paper we exploit the predecessor/successor relation between the train tasks.

Both rolling stock and crew operate according to plans which are detailed to a daily level i.e. for each train task it is known which specific driver and which specific train units will cover the train task. The rolling stock and crew plans are assumed optimal for the situation without disturbances. Therefore, given a disturbance to either of the plans, we seek to return to the original plan as soon as possible. Returning to the original plan means that each train unit returns to its originally planned path, which eventually will route the train unit to the maintenance center.

A set of trains with the same stopping pattern and a uniform frequency between

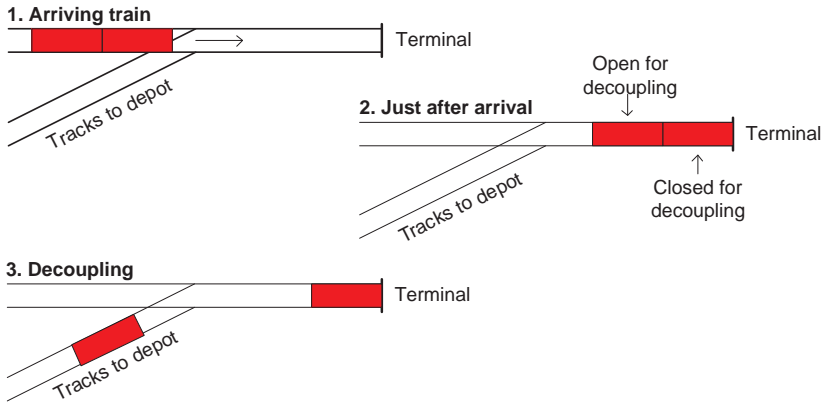


Figure E.3: Illustration of the open and closed end of a composition at the terminal station

trains form a *Train line*. The train line concept is first of all used externally for representing the timetable to the customers, but it is also used internally for planning and prioritizing.

A rolling stock schedule consists of a set of *Train unit routes* where each route refers to a specific train unit and covers a path of train tasks. These train tasks may or may not belong to the same train sequence.

When a train unit leaves or is added to a train sequence it is said to be decoupled from or coupled to the train task. The set of train units assigned to a train is called a *composition*. As mentioned earlier, the composition defines the number of each type of train units and the order in which they are coupled. At S-tog there are two different train unit types. These can be coupled in all possible combinations limited by a maximum length of the train.

At S-tog coupling/decoupling always occurs at only one end of the train depending on the depot at which the coupling/decoupling occurs i.e. the train is only open for coupling/decoupling in one end. The route of a train unit must be feasible with respect to the open end of the train. That is, if a train unit is to be decoupled from a train, it must be in the open end of the composition. When coupling a train unit to a train, the train unit must also be assigned to the open end of the train. The open versus the closed end of a composition at a terminal is illustrated in Figure E.3.

E.4 Defining a disruption

Incidents occur in real time that disturb the planned operation. Some of these incidents are of such a size that also the rolling stock plan is disturbed. For a more detailed description of the effect disruptions have on the S-tog timetable see Hofman et al. [2006].

To minimize the impact of an incident, network controllers employed by the infrastructure owner reroute trains to get operation back to normal as quickly as possible.

The delays disturbing the timetable may, as mentioned, be of a size that also disrupts the rolling stock and crew schedules.

A rolling stock schedule is disrupted when train units are not able to cover the train tasks they were expected to cover. The rolling stock schedule is affected by the delays both directly and indirectly. An example of a directly disrupting effect is the break down of a train unit thereby causing the train unit not to be able to cover its scheduled train task. Indirectly, the rolling stock schedule is affected by the actions of the train route dispatchers trying to return the departures to normal.

There are several potential negative consequences of a disruption in the rolling stock schedule. A rolling stock disruption may imply an imbalance in the rolling stock available at the rolling stock depots. This again may lead to train tasks being insufficiently covered according to their expected passenger demand. Another secondary disruptive effect can be that the reallocation of train units to train tasks other than the originally scheduled ones may lead to broken maintenance constraints for individual train units.

The set of train units being assigned extraordinarily to cover another train sequence are not necessarily of the same type and number as the set of train units originally intended for that train sequence. Hence, future couplings/decouplings on the train sequence and other trains running on the same route may also be affected.

E.4.1 Objectives when minimizing rolling stock disruption

The rolling stock dispatcher does not have the time to take into account several objectives when minimizing the extent of a disruption to the rolling stock plan.

He tries to minimize the number of departures not covered and chooses the first feasible solution he discovers in the manual solution process.

Several objectives are interesting to include in a rolling stock recovery model. Fioole et al. [2006] mention seat shortage, efficiency and robustness as relevant for the operational planning phase. These are also relevant in real time.

Seat shortage refers to the difference between the number of seats on the train units allocated to a train task and the expected seat demand of the train task. Maximizing the efficiency means that we do not want to operate a train task with more train units assigned than necessary, either considering the number of excess seats or the number of train unit kilometers driven. The two objectives of seat shortage and efficiency can be conflicting and will hence have to be weighted. Robustness in a rolling stock recovery plan is translated directly to the number of couplings and decouplings planned in a recovery plan. A recovery plan with many couplings and decouplings is less robust than one that has fewer. We wish to maximize robustness in a plan given that we still weigh the objectives of seat shortage and efficiency against each other. Robustness is therefore also assigned a weight in the final objective function.

Seat shortage, efficiency and robustness are all objectives concerning the assignment of train unit types to train tasks. Other objectives concern the physical train units. In real time the aim is to recover to the original rolling stock plan. However, it may not be possible within the time window of recovery or even within the same day of operation to route the train units back to their original work plans. Hence, an objective to include in the objective is the difference in end depot balance between the original and the recovered plan.

E.4.2 Basic concepts in a disruption

It is likely that several delays call for recovery occur during the day. In the real-time situation time is a critical factor and recovery decisions must be made fast. For each recovery scenario we therefore solve within a specified time window e.g. two hours and include a limited set of train units. The start and end of the time window is the considered start and end time of the disruption.

Typically, all train units, $k \in K$, assigned to the train lines of the affected train units are included in the recovery scenario plus possibly some of the other lines running on the same route and sharing the same depots. Also, all train units located on the affected depots at the start time of disruption will be included in the set of train units to be replanned for.

Type	Length	No Seats
SE	46	150
SA	86	336

Table E.1: rolling stock types

Each train unit has a kilometer limit, $KmLimit_k$. It indicates the maximum number of kilometers that the distances of the tasks assigned to the train unit during recovery must sum up to. Each train unit has a seat capacity matching its train type. For each train unit the start depot, $\delta_\alpha(k)$, and a preferred end depot, $\delta_\omega(k)$, are given.

At all times two rolling stock types, $m \in M$, are available. These are short and long train units named SE and SA respectively. Sizes of the two rolling stock types are listed in Table E.1.

The train tasks, $t \in T$, considered are those left uncovered, those which are insufficiently covered w.r.t. demand and those for which the assigned train units have been included in the recovery scenario.

For each train task, t , the start and end time, $\tau_d(t)$ and $\tau_a(t)$, and start and end depot, $\delta_d(t)$ and $\delta_a(t)$, are known. Each t is associated with a length in kilometers, Km_t , and a duration measured in seconds, $Time_t$. The set of tasks having no predecessors constitutes T_0 . The train tasks having no successors constitute T_1 . The successor of the train task t is denoted $\nu(t)$. Each train task has a seat demand, $Demand_t$.

The set of depots involved in the recovery scenario, D , is defined by the routes of the train lines included. For each depot, $d \in D$, included in the recovery scenario the start capacity of each type of train unit m is given by $DepotCap_{d,m}$.

Composition	Order	NO seats	Length
0	SE	150	46
1	SE - SE	300	92
2	SA	336	86
3	SE - SA	486	132
4	SA - SE	486	132
5	SA - SA	672	172

Table E.2: Compositions

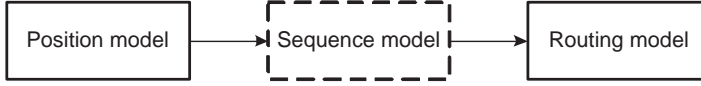


Figure E.4: The decomposed approach for the RSRP. The dotted line illustrate that the Sequence model is optional in the solution process.

The maximum length of the composition assigned to a train is equivalent to the length of two SA train units. Given this maximum length, in fact a composition consisting of three SE train units or a composition consisting of two SE and one SA train unit are applicable in practice. Even though these train composition consisting of three train units are feasible, we omit them from our model. Seen from a modelling perspective our model is significantly reduced in size when reducing the number of allowed train units from three to two. Seen from a practical perspective, only few train tasks at S-tog will normally be assigned three train units. More specifically, at the tactical planning level no train tasks will be assigned more than two units. In a recovery situation three units on a train task occurs not even on a daily basis.

E.5 Network description

We address the Rolling Stock Recovery Problem (RSRP) through a decomposition approach as we evaluate the problem too comprehensive to be solved in one step. An initial model (the Position model) decides the number and order of train units of each train type for each train task. Reformulated, the Position model decides the composition for each train task. Secondly, the specific train units are routed according to the solution of the initial model. The second step is decomposed into two models each dealing with the problem of assigning train units to train tasks, but at different aggregation levels.

The first routing model (the Sequence model) addresses the routing of train units by considering train tasks in sets of train sequences. The second routing model (the Routing model) addresses the train tasks individually. The interaction of the three models is illustrated in Fig. E.4. All three models are assignment models with side constraints where balances of commodities respectively feasibility of flow are handled indirectly. The models are described in Sections E.6 to E.8.

The objectives mentioned in Section E.4.1 are all included in the Position model

where the number and order of train units of each train type is decided for each train task.

E.6 The train unit position model

In this section we introduce the variables, objective, and constraints of the Train Unit Position model (the Position model). The main variables of the model describe the assignment of train type to train task and position.

$$X_{tp}^m = \begin{cases} 1 & \text{If a train unit of type } m \text{ is assigned to task } t \text{ in position } p \\ 0 & \text{Otherwise} \end{cases}$$

From these X -variables the L -variables are derived. The L_t^m variables are inventory variables indicating the number of train units of type m present at the departure depot of t immediately before the departure of t .

Finally, O_t^m and N_t^m are variables indicating whether respectively coupling and decoupling is carried out between the tasks t and $\nu(t)$. Both sets of variables are binary.

L^0 are the start inventory parameters. L_{dm}^0 indicates the number of train units of type m located in depot d at the beginning of the disruption. L_{dm}^1 are the end capacity variables indicating the number of train units of type m present at depot d in the end of the considered recovery period. A desired end depot capacity is given by the parameter $E[cap]_d^m$. The variables E_d^m indicate the shortage of train units of type m in depot d in the end of the recovery period.

L_t^m are calculated from L_{dm}^0 and X_{tp}^m . As both are integers, the L -variables will automatically be integer. Therefore, we only require that $L_t^m \in \mathbb{R}_+$, $\forall t \in T, m \in M$.

The relevant aspects we include in the objective of the positioning model are seat shortage, number of composition changes, the cost of covering train tasks with train units and the sum of differences to the originally scheduled capacity on the depots, see Eq. E.1.

$$\begin{aligned}
\text{Minimize OBJ} = & \\
& W_1 \cdot \sum_{t \in T} (\text{Demand}_t - \sum_{m \in M, p \in P} \text{Seats}^m \cdot X_{tp}^m) + \\
& W_2 \cdot \sum_{t \in T, p \in P, m \in M} K m_t \cdot X_{tp}^m + \\
& W_3 \cdot \sum_{t \in T} (\sum_{m \in M, p \in P} \text{Seats}^m - \text{Demand}_t \cdot X_{tp}^m) + \\
& W_4 \cdot \sum_{t \in T, m \in M} O_t^m + W_5 \cdot \sum_{d \in D, m \in M} E_d^m
\end{aligned} \tag{E.1}$$

As a train has a maximum length each train task cannot be covered by more than the maximum number of train units per train. This is guaranteed by Eq. E.2.

$$1 \leq \sum_{m \in M, p \in P} X_{tp}^m \leq \text{MaxTrainLength}, \quad \forall \quad t \in T \tag{E.2}$$

Physically at most one train unit can be assigned to each position of a train task. Eq. E.3 ensures this.

$$\sum_{m \in M} X_{tp}^m \leq 1, \quad \forall \quad t \in T, p \in P \tag{E.3}$$

We control the incoming and outgoing flow of depots by three sets of inventory constraints, see Eq. E.4 to E.6.

The first set of inventory constraints controls that the initial inventory level is not violated. This means that for each depot d the tasks departing before the first arriving task can not use more capacity than what is present initially given by L_{dm}^0 . The set of departing tasks before the first arrival task on depot d is denoted ϕ_d for all $d \in D$. See Eq. E.4.

$$\sum_{p \in P, t \in \phi_d} X_{tp}^m \leq L_{dm}^0, \quad \forall \quad d \in D, m \in M \tag{E.4}$$

The inventory in a depot of train unit type m immediately after the arrival of a train task t is given by the start capacity on the depot minus the sum of every train unit of type m coupled to train tasks at that depot before and including t and plus the sum of every train unit decoupled from train tasks at that depot before and including t . This is handled by Eq. E.5.

$$\begin{aligned}
L_t^m = L_{\delta_a(t)m}^0 - \sum_{\substack{p \in P, t' \in T \\ \tau_d(t') \leq \tau_a(t) \\ \delta_d(t') = \delta_a(t)}} X_{t'p}^m + \\
\sum_{\substack{p \in P, t' \in T \\ \tau_a(t') \leq \tau_a(t) \\ \delta_a(t') = \delta_a(t)}} X_{t'p}^m, \quad \forall \quad t \in T, m \in M
\end{aligned} \tag{E.5}$$

The last set of inventory constraints concerns the end capacity. The end capacity, L_{dm}^1 , of train unit type m in depot d is given by L_t^m for which t is the last traintask arriving on d , θ_d . See Eq. E.6.

$$L_{dm}^1 = L_{\theta_d}^m, \quad \forall \quad d \in D, m \in M \tag{E.6}$$

We wish to control the end depot balance by minimizing in the objective function the shortage of train units defined by variables E_d^m . These are defined in Eq. E.7

$$E_d^m \geq E[cap]_d^m - L_{dm}^1, \quad \forall \quad d \in D, m \in M \tag{E.7}$$

Each depot has an individual upper capacity on the number of units that can be stored at that depot. The upper capacity is estimated by controlling the length of the rolling stock stored at each depot relative to the length of the depot tracks, $DepotCap_d$. Eq. E.8 controls the capacity of each depot right after the departure of each task, that is, $\delta_d(t)$ is the departing depot of t .

$$0 \leq \sum_m Length_m \cdot L_t^m \leq DepotCap_{\delta_d(t)}, \quad \forall \quad t \in T \tag{E.8}$$

The coupling and decoupling variables are determined in Eq. E.9 and E.10. We use a constant B to find the O_t^m and N_t^m variables. This is potentially very expensive considering computation time when B has a high value, however, B can be limited to the maximum train length plus one and as the maximum train length is 2 units B has a low value.

$$B \cdot O_t^m \geq L_{\nu(t)}^m - L_t^m, \quad \forall \quad t \in T \setminus T^1, m \in M \tag{E.9}$$

$$B \cdot N_t^m \geq L_t^m - L_{\nu(t)}^m, \quad \forall \quad t \in T \setminus T^1, m \in M \tag{E.10}$$

To ensure that no train unit is decoupled from a train if it is positioned in the closed end of the train composition, we use one of the set of equations in Eq. E.11 depending on the value of the 0-1 parameter $ChangePosition_t$. This parameter indicates whether open position is changed from one end of the train to the other after train task t .

$$ChangePosition_t = \begin{cases} 1 & \text{If closed position of task } t \text{ is different from} \\ & \text{closed position of successor } \nu(t) \\ 0 & \text{Otherwise} \end{cases}$$

$$\begin{aligned} & \text{If } ChangePosition_t = 1 \\ & \quad X_{tp}^m \leq \sum_{p' \in P, p' \neq p} X_{\nu(t)p'}^m + W_{\nu(t)} \\ & \quad X_{tp}^m \leq \sum_{p' \in P, p' = p} X_{\nu(t)p'}^m + V_{\nu(t)} \\ & \text{else} \\ & \quad X_{tp}^m \leq X_{\nu(t)p}^m \end{aligned} \quad (E.11)$$

V_t and W_t are length indicator variables. Both sets of variables are binary. V_t is one if one train unit is assigned to task t and zero otherwise for all $t \in T$. W_t is one if two train units are assigned to task t and zero otherwise for all $t \in T$. They are determined through Eq. E.12 and E.13.

$$W_t \leq \sum_{p \in P, m \in M} X_{tp}^m - 1 \leq 2 \cdot W_t, \quad \forall \quad t \in T \quad (E.12)$$

$$V_t + W_t = 1, \quad \forall \quad t \in T \quad (E.13)$$

The results achieved when solving the TUP model are comparable to the results that are achieved when solving the model described in Fioole et al. [2006]. The difference between the two models lies in the handling of the compositions. In the Fioole model the compositions are handled as a set of train unit types i.e. a composition is assigned to each train task and binary variables describe specifically the transformation from composition to composition on consecutive train tasks on the same train sequence. In our model we handle the positions in the train's composition specifically. The Position model is a feasible choice due to that the maximum length of compositions on train tasks is limited to 2 train units.

E.6.1 Size of model

Given a time window of the disruption of two hours and including all train lines intersecting the most dense part of the S-tog network, a total of 550 train tasks result. We have restricted the problem to only consider compositions up to two unit as opposed to the real restriction of three units. The Train unit position model has therefore approx. 6500 variables and approx. 6500 constraints. An expanded model for the problem considering compositions of up to three train units (the S-tog maximum length of composition) will have approx. 8000 variables and approx. 8500 constraints. This is an estimate as Equation E.3 must be changed according to the new maximum composition length.

In comparison the Fiole model in comparison has approx. 27000 variables and 12000 constraints when considering compositions up to two train units and approx. 41000 variables and 16000 constraints when considering compositions up to three train units.

E.6.2 Solution approach for Train unit position model

The model is implemented in C# using Concert Technology from ILOG and solved using Cplex 10.0. Given the size of the problem we expect solutions to be achieved within acceptable computation times.

E.7 Train sequence model

When a train unit's path consists of one train sequence it is certain that the train unit is not decoupled or coupled at any time. That is, coupling and decoupling refer to train unit flows to and from the train sequence. Both are time demanding and in a periodic timetable there will not necessarily be sufficient time for performing these. It is assumed that if the number of couplings/decouplings are decreased the robustness of the rolling stock plan is increased. That is, the rolling stock plan will be less sensitive towards minor interferences in the operation.

This section describes the Train sequence model (Sequence model). The model is an assignment model, which if possible assigns a single, physical train unit to each train sequence in the disruption scenario, such that the train unit can

feasibly cover the entire train sequence. In this way the model mirrors qualities that are part of solutions known to work well in practice.

The consequence of only covering the set of train sequences with one train unit each (if in fact a train unit exists that can make a feasible cover) is that for a set of train tasks the demand will not be fully covered. For some of the train tasks one train unit will be assigned but the demand exceeds the seat capacity of that train unit. For some train tasks no train unit will be assigned and the demand not covered at all. The set of train tasks not covered sufficiently will be addressed in a third model.

There is a preference of which train unit type to assign in the process of assigning train units to train sequences. The preferred train unit type is chosen given the results from the Train unit position model. Recall that this model gives information regarding number and type of train unit types assigned to each train task. For each train sequence the train unit type chosen as the preferable coverage is the type being present on each composition of the train tasks of the train sequence.

The Train sequence model has one set of variables, ϕ_s^k , which assign physical train units to train sequences.

$$\phi_s^k = \begin{cases} 1 & \text{If train unit } k \text{ is assigned to train sequence } s \\ 0 & \text{Otherwise} \end{cases}$$

The objective function of the Train sequence model is to maximize the sum of preferences of train units, k , assigned to train sequences, s , see Eq. E.14. As many train sequences are assigned a train unit as possible provided that a train unit exists for the train sequence that contributes to a feasible solution. The preference of assigning train unit k to train sequence s is c_s^k . It takes the value of 1 if train unit k is a possible match for sequence s and -1 if it is not.

$$\text{Maximize} \quad \sum_{s \in S, k \in K} c_s^k \cdot \phi_s^k \quad (\text{E.14})$$

Each train can be covered by at most one train unit. The train unit, k , must have the same start and end depot, $\delta_\alpha(k)$ and $\delta_\omega(k)$, as the train sequence. Start and end depot of the train sequence s are denoted $\delta_\alpha(s)$ and $\delta_\omega(s)$. This is ensured by Eq. E.15.

$$\sum_{\substack{k \in K, \\ \delta_\alpha(s) = \delta_\alpha(k) \\ \delta_\omega(s) = \delta_\omega(k)}} \phi_s^k \leq 1, \quad \forall s \in S \quad (\text{E.15})$$

For the train unit covering a train sequence maintenance requirements must be respected. This is easily included in the Sequence model, see Eq. E.16. $EndRun_s$ is a parameter indicating the number of kilometers which are left after recovery until the depot is reached. This can be derived from the original rolling stock plan. $KmBefore_k$ is a parameter indicating the number of kilometers that the unit has driven before the start of the recovery plan.

$$\sum_{s \in S} (Km_s + EndRun_s) \cdot \phi_s^k \leq KmLimit_k - KmBefore_k, \quad \forall k \in K \quad (\text{E.16})$$

The 550 train tasks mentioned in the dimensioning of the Train unit position model groups into less than 70 train sequences. Available for covering the problem are at most 130 train units. This results in approximately 9000 variables and less than 350 constraints.

Again the model is not of considerable size and we solve it using Cplex 10.0 and Concert Technology where the model is implemented in C#.

E.8 The train Routing model

As mentioned in the previous section E.7 the Train sequence model will only cover some of the train tasks according to their respective demands. Some will either be left uncovered or covered insufficiently according to demand. These must be covered by valid train task paths using the train units not yet assigned to a train path. This is done by the Train Routing Model, which is an assignment model considering each train task individually.

The main variables of the Train Routing model are, q_t^k . These variables assign train units to train tasks.

$$q_t^k = \begin{cases} 1 & \text{If train unit } k \text{ is assigned to train task } t \\ 0 & \text{Otherwise} \end{cases}$$

To control the solutions of the model a second set of variables is introduced, ρ_t^k .

The ρ_t^k variables are used to control the number of couplings/decouplings in the solution.

$$\rho_t^k = \begin{cases} 1 & \text{If train unit } k \text{ is assigned to train task } t \text{ and to the successor of } t, \nu(t) \\ 0 & \text{Otherwise} \end{cases}$$

A set of artificial tasks are added to the problem representing the sources, T_{so} , and sinks, T_{si} , of train tasks. There are $|K|$ sources and $|D| \times |K|$ sinks. The set of train tasks are in the set T_{tasks} . The joint set of tasks is $T = T_{tasks} \cup T_{so} \cup T_{si}$.

The objective function maximizes the total sum of covered demand and the sum of couples of consecutive tasks covered by the same train unit. The use of physical train units also included in the objective by the sum of sources and sinks. All terms are weighted using weights, W_1 to W_4 . See Eq. E.17.

$$\begin{aligned} \text{Maximize} \quad & W_1 \cdot \sum_{t \in T_{tasks}, k \in K} q_t^k + W_2 \cdot \sum_{t \in T_{so}, k \in K} q_t^k + \\ & W_3 \cdot \sum_{t \in T_{si}, k \in K} q_t^k + W_4 \cdot \sum_{t \in T_{tasks}, k \in K} \rho_t^k \end{aligned} \quad (\text{E.17})$$

Each train task must be covered at most corresponding to the number of each train unit type assigned to the task in the Position model, see Eq. E.18 and E.19. The parameter $cars_m$ represent the number of cars on train unit type m . constraining the number of cars and the number of train units on a train task to be the same in the Routing model as in the Position model, we are ensured that the right train composition is assigned to the train task.

$$\sum_{k \in K, type_k = m} cars_{m(k)} \cdot q_t^k \leq \sum_{p \in P, m \in M} cars_m \cdot X_t^{m,p}, \quad \forall \quad t \in T_{tasks} \quad (\text{E.18})$$

$$\sum_{k \in K, type_k = m} q_t^k \leq \sum_{p \in P, m \in M} X_t^{m,p}, \quad \forall \quad t \in T_{tasks} \quad (\text{E.19})$$

The ρ_t^k variables are defined in Eq. E.20.

$$2 \cdot \rho_t^k \leq q_t^k + q_{\nu(t)}^k, \quad \forall \quad k \in K, t \in T_{tasks} \setminus T^1 \quad (\text{E.20})$$

The train tasks assigned to a train unit must form a valid train route i.e. a path through the network, which is feasible with respect to time and place of

each adjacent pair of train tasks on the route. Also, the train route for each individual train unit must be valid with respect to any required start and end depots of the train unit. We add a set of virtual nodes to the network, one set representing the source nodes, N_{so} , of each individual train unit and one set representing the sink nodes, N_{si} , of each individual train unit. For each train unit there is a sink node for each depot i.e. there are $|Depots| \cdot |Trainunits|$ sinks in total.

The constraints ensuring valid paths are in Eq. E.21 to E.25. Eq. E.21 ensures that if the source of a train unit is not covered, the train unit is not covering any of the train tasks. Eq. E.22 ensures that if the source is covered for a train unit, then so is exactly one of the sinks of the train unit. Eq. E.23 and E.24 are equivalent to the flow constraints of a multi commodity flow model. They ensure that if train unit k is covering train task t then at least one of the predecessors, $pred(t)$, respectively successors, $succ(t)$ are covered. Finally, Eq. E.25 ensure that if train unit k is assigned to t then it can cover none of the train tasks parallel in time to t . Time parallelism is illustrated in Fig. E.5. The four tasks t_1 to t_4 are all time parallel to t because they intersect the time interval between departure time and arrival time of t . The parameter n in Eq. E.25 indicates the maximum number of train tasks present within the time interval of t on any other sequence in the relevant problem instance. See Fig. E.6.

$$q_t^k \leq q_{t'}^k, \quad \forall \quad k \in K, t' \in T_{source(k)}, t \in T_{tasks} \setminus T_{source(k)} \quad (E.21)$$

$$\sum_{t \in T_{sinks(k)}} q_t^k - q_{t'}^k = 0, \quad \forall \quad k \in K, t' \in T_{source(k)} \quad (E.22)$$

$$\sum_{t' \in T_{pred(t)}} q_{t'}^k \geq q_t^k, \quad \forall \quad k \in K, t \in T_{tasks} \quad (E.23)$$

$$\sum_{t' \in T_{succ(t)}} q_{t'}^k \geq q_t^k, \quad \forall \quad k \in K, t \in T_{tasks} \quad (E.24)$$

$$\sum_{t' \in T_{parallel(t)}} q_{t'}^k \leq n - (n - 1) \cdot q_t^k, \quad \forall \quad k \in K, t \in T_{tasks} \quad (E.25)$$

Note that the Train position model and the Train ID model can function without the Train Sequence model. The Train sequence gives us two advantages when included in the solution process. First, it heavily reduces the number of variables that must be taken into account in the Train routing model. Second,

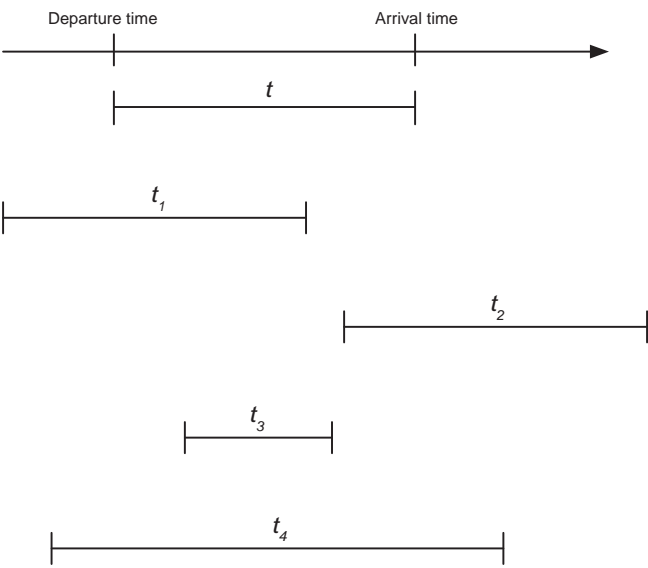


Figure E.5: Illustrating time parallelism: t_1, \dots, t_4 are all time parallel with t

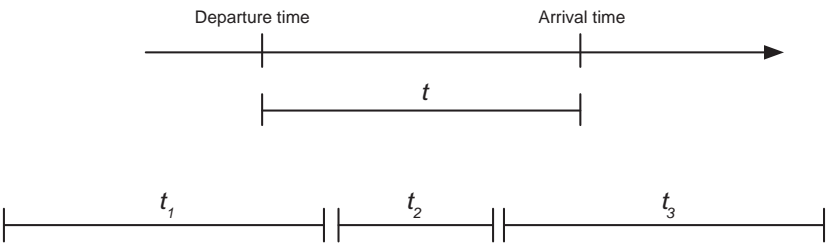


Figure E.6: Illustrating the meaning of parameter n : Three train tasks are present in the train sequence during the time span of t

we decrease the number of broken composition constraints. The disadvantage is that decomposing into three models instead of two may give a solution farther from optimal. However, the train sequence model imitates features of solutions working well in practice. When the train sequence model is included in the solution process the constraints in Eq. E.26 are included in the train routing model ensuring that train unit k is assigned to train task t if t is in train sequence s and s has been covered by k in the Sequence model.

$$\Phi_s^k = 1 \Rightarrow q_t^k = 1, \quad \forall \quad k \in K, t \in T_{tasks} \quad (\text{E.26})$$

We implement the model using Concert Technology and solve the model with Cplex. There are, however, potentially more than 75,000 variables and solving the model with Cplex is expected to be too time consuming. The large number of variables stem mainly from the q_t^k variables, which account for 60,500 of the total. The rest are the auxiliary variables.

E.9 Computational results

Extensive experiments have been carried out for the decomposed approach. We first discuss experiments with the main purpose of choosing a setting of the weights in the objective function of the Position model. The weights must provide a sufficiently good solution quality and a sufficiently short computation time. The second set of experiments aims at determining a weight setting for the Routing model objective function. Finally, we present experiments illustrating the different results achieved when respectively including and excluding the Sequence model in the solution approach.

E.9.1 Experimental results for Position model

A set of experiments on various weight settings for the objective function of the Position model form the basis for further experiments. The aim of the experiments is to derive a set of weights for which solution quality and computational time are both acceptable. The experiments will be constructed as a statistical design of experiments (DOE), see Montgomery [1997].

Two sets of factor experiments is conducted each having a statistical DOE. In the first set of experiments six factors of varying levels are included, see Tab. E.3.

Factor	Type
A	Standing passengers
B	Train unit kilometers
C	Excess seat
D	Composition Changes
E	End capacity difference
F	Instance size

Table E.3: The factors with varying weights.

Level	A	B	C	D	E	F
0	0	0	0	0	0	C
1	1	1	1	1	10^4	A, A+
2	10		10	1000	10^5	
3	100		100			

Table E.4: The different levels used for factor experiments.

The second set of experiments includes factors A to E. Factor A to E represent the weights of the objective function of the Position model as described in Tab. E.3.

We have used the values presented in Tab. E.4 for each weight.

A full design of experiments contains 288 instances without factor F and 576 including factor F. We have used a design limiting the number of experiments to 72 for all experiments where each of the 72 experiments is equivalent to a specific weight setting of the objective function.

The disruptions are based on real-life data from the timetable in 2006. A data set is chosen with low punctuality and in which train units ended up in wrong locations according to their individually planned end station. A disruption is limited within a time window. The train tasks included in the disruption intersect the time window and are included in the train sequences of a set of train lines given as input. The train units included in the disruption are those being assigned to the included train tasks plus the train units being located at the depots of the train lines at the start of the disruption time window.

We run two types of experiments. In the first type, factor F is included at the levels shown in Tab. E.4. In the second type we exclude factor F. We have run

4 different sets of lines, $A\&A+$, C , E and $H\&H+$ for the type 2 experiments¹. The experiments were run with an upper limit on the solution time on 300 seconds.

We have used the method described in Nielsen and Christensen [2006], to develop the DOE. A statistical function for a general linear model is derived using information from the 72 experiments. The function can be used for estimating the contribution of each factor to the objective for some parameter setting. We assume that the contributions from third order correlations and higher are negligible. For a DOE with three factors, the statistical function is shown in Eq. E.27. A function for 6 factors A to F follows the same structure.

$$F_{OBJ} = A + B + C + AB + AC + BC + \varepsilon \quad (\text{E.27})$$

The basic idea of using DOE is to reduce the number of experiments necessary to gain information on the contribution and importance of each term in the objective. By calculating the value of the statistical objective function and comparing it to the values observed in the results, we get an impression of how well the chosen experiments describe the effect from each factor. If the average error, ε , is low the chosen experiments are assumed representable for choosing a weight set for the objective function of the mathematical model that can be used in further experiments. We also evaluate the contributions from each factor on each of the terms in the objective. If the contribution is as expected, we assume the experiments representable and thereby a sufficient basis for choosing a weight set for the objective function of the mathematical model that can be used in further experiments.

We use the statistical function to calculate the contribution of factor A to E, the computational time and the joint objective function. In Tab. E.5 the average error contributions measured in percentage of the average observation are listed for all experiments.

***Type*₁ experiments:** For experiments of *Type*₁ we see that the average error contributions for all terms are lower than those of *Type*₂ except for factor A. The low error contributions indicate that the *Type*₁ experiments are representative, however, evaluating the contributions from each factor on all terms in the objective we observe that the contributions cannot be reasonably explained. For example, factor C at the high levels contributes to the kilometer term of the objective function of the Position model, see Appendix E.A. This is a contradiction as factor C relates to the standing passengers. If the number of standing passengers are decreased by the model more train units are used and hence the

¹The S-tog lines are illustrated in the S-tog network in Fig. E.7



Figure E.7: The S-tog network

Factor	ε_{Type_1}	$\varepsilon_{Type_2}^{A\&A^+}$	$\varepsilon_{Type_2}^C$	$\varepsilon_{Type_2}^E$	$\varepsilon_{Type_2}^{H\&H^+}$
A	7.93	6.05	6.05	8.75	4.50
B	1.50	3.71	3.09	2.47	1.46
C	1.29	11.90	10.44	8.61	4.90
D	2.23	9.32	4.48	4.41	5.16
E	1.32	27.91	64.08	14.40	9.82

Table E.5: The average error contribution for the different terms in the objectives given the estimated objective function.

number of train unit kilometers is increased. Another example of a contradiction is that factor C punish itself at all levels. The first order factor contributions are enclosed in Appendix E.A.

Because of the lack of consistency between expected and actual contributions we conclude that $Type_1$ experiments are not representative.

$Type_2$ experiments: Considering the $Type_2$ experiments we make the following observations on the error contributions and the first order factor contributions²:

$A\&A^+$: The error contributions are especially high for factor E and C. Also, if we consider the different contributions that the factors make to each term in the objective there are contradictions similar to the ones observed for $Type_1$ results.

C : The error contributions are especially high for factor E and C. The contributions from each factor on each term in the objective are all as expected.

$E\text{and } H\&H^+$: For both the experiments on E and $H\&H^+$ the error contributions are especially high for factor E. The contributions from each factor on each term in the objective are all as expected.

For all line combinations but $A\&A^+$ the instances solve to optimality within the computation time limit of 300 seconds. A large part of the $A\&A^+$ instances do not find the optimal solution within the 300 seconds. The error contribution and the lack of ability to describe the contributions of the $A\&A^+$ instances indicate that these are not representative. As they are not representative, we will not use them for determining the weight set used for further experiments.

²The first order factor contributions are enclosed in Appendix E.B

Given the average error contributions in Tab. E.5 and the evaluation of the expected versus the actual contributions commented above, we base our choice of a weight setting for further experiments on the instances of C , E and $H\&H+$.

We also investigated whether one can trace dependency between the computational time and the weight setting used for the objective. However, results show that there is no connection. When we use the statistical function for estimating the computational time the average error contribution varies from 20 to 75 %.

Choice and validation of weight setting

We have chosen the set of weights by filtrating the experimental results with respect to the criteria listed below.

1. Choose a subset of instances with lowest end capacity difference.
2. Choose a subset of instances where the maximum number of standing passengers is low. Preferably the maximum number of standing passenger should not exceed 36. This is 10 percent of the seat capacity in an SA train unit.
3. Choose the experiments which has the lowest average values of excess seats.
4. Choose the set of instances with lowest number of driven train unit kilometers.

Given a selection of instances, which are based on the criteria above, we assume that results are satisfactory with respect to all terms in the objective function. Based on the sorting and filtration we have chosen the instance that has a short computation time. The final choice of weight setting used for all further experiments is the combination $Weights_1 = (100, 1, 10, 0, 10^4)$. Furthermore we have chosen one more weight set, $Weights_2 = (1, 1, 1, 100, 10^5)$, for comparison. $Weights_1 = (W_1, W_2, W_3, W_4, W_5)_1$ and $Weights_2 = (W_1, W_2, W_3, W_4, W_5)_2$ are parameters used for the objective function in the Position model where W_i is the weight on the i th term in the objective function. We expect that $Weights_1$ emphasizes specifically the number of standing passenger whereas we expect that $Weights_2$ puts a higher emphasis on number of driven kilometers and the amount of excess seats, though standing passengers are still given some importance.

Lines
A
C
E
H
A+
H, H+
A, A+
C, H+
H, C
C, A+
E, A+
E, A
E, A, A+
C, H, H+
E, H+, C
E, A+, C
E, H+, A
E, C, A+, A
E, C, H, H+
H, H+, C, A+, A
H, H+, C, A+, A, E

Table E.6: Lines included in experiments, see Fig. E.7.

We have run a set of experiments on each of the two weight sets. The purpose of the experiments is to verify the expected difference of objectives for each of the two weight settings and to see if $Weights_1$ are more likely to have a short computation time than $Weights_2$. Each experiment is defined by a set of lines and recovery time window. The line sets are represented in Tab. E.6. The recovery windows are respectively 1, 2 and 3 hours in the morning peak hour starting from 7 o'clock. The line combinations listed in Tab. E.6 combined with the three different time periods gives 63 experimental instances. As explained these instances are run for two weight sets which gives a total of 126 experiments. The upper limit on computational time for each instance is 3600 seconds.

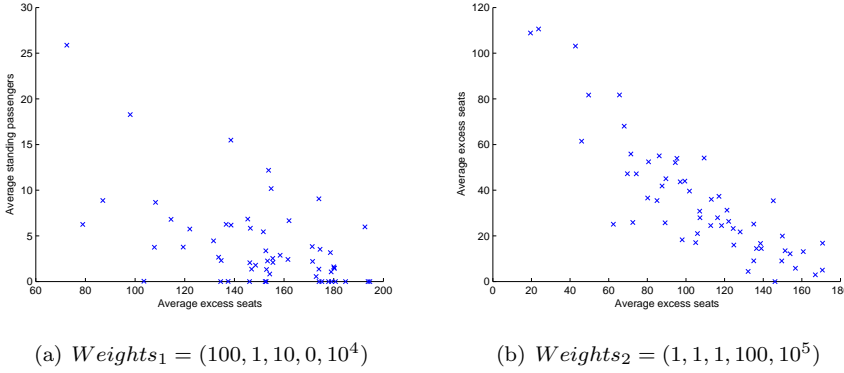


Figure E.8: Each point in the plots is the average for a solution over all its train tasks of number of respectively standing passenger and excess seats

In both plots in Fig. E.8 the average excess seats versus the average standing passengers for each of the 63 experimental instances are illustrated. Each point in the plots relates to a solution. Notice that it is possible in a solution to have both an average number of excess seats and an average number of standing passengers larger than zero as we average over all train tasks in that solution. For a single train task the number of respectively standing passengers and excess seats cannot both exceed zero.

If we inspect the two figures in E.8 we see that the instances illustrated in E.8(a) as expected in general have much fewer standing passengers on average than the instances in E.8(b). The average numbers of excess seats in the $Weights_1$ solutions are not much higher than numbers of excess seats in the $Weights_2$ solutions. For both figures the relationship between the average number of standing passengers and the average number of excess seats seems approximately linear.

In Fig. E.9 the two plots show the sum of excess seats versus the number of composition changes for each experimental instance. The numbers of composition changes only vary little from the $Weights_1$ solutions to the $Weights_2$ solutions. Both Fig. E.9(a) and E.9(b) indicate a linear relationship between composition changes and excess seats.

The two plots in Fig. E.10 shows the sum of standing passengers versus the number of end capacity differences for each experimental instance. There are much fewer standing passengers in the $Weights_1$ solutions, see Fig. E.10(a) than in the $Weights_2$ solutions, see Fig. E.10(b). The number of depot end capacity

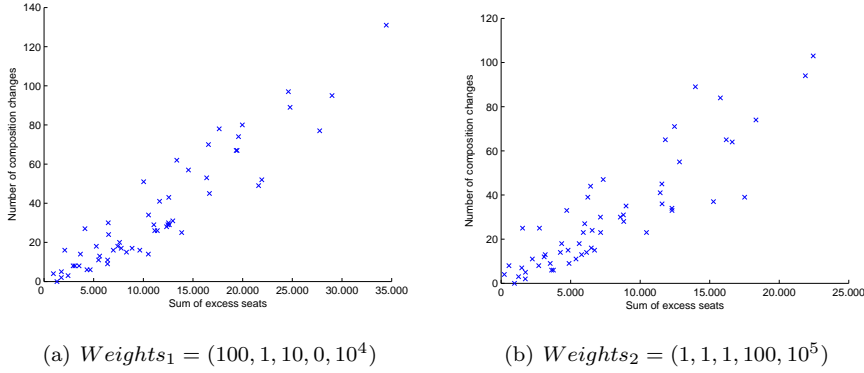


Figure E.9: Sum of excess seats versus number of composition changes.

differences only vary little, however, a tendency shows that a high emphasis on few standing passengers results in relatively more end capacity differences. The number of end capacity differences do not increase much in the $Weights_1$ solutions.

We have chosen $Weights_1$ partly because these weights lead to low computation time. We are interested in whether the low computation time observed in the initial experiments is low in general. We therefore compare computation times of $Weights_1$ results with those of $Weights_2$. In Fig. E.11 the differences in solution time between the $Weights_1$ solutions and the $Weights_2$ solutions are illustrated. Generally there is little difference between the solution time for the two weight sets, however, there is a set of 8 to 10 problems that solve in much shorter time for $Weights_1$. Solution times for $Weights_1$ are approximately 90% faster than those of $Weights_2$. There is only one instance where $Weights_2$ is much faster than $Weights_1$.

General comments on Position model results

There is a large variation in the results of the experiments with respect to depot end capacities. The quality with respect to standing passengers and excess seats varies independently of the depot end capacities. Finding a good balance between standing passengers and excess seats may effect the depot end capacities. A high weight on depot end capacity will often increase both the number of standing passengers and the excess seats. When we assign a low

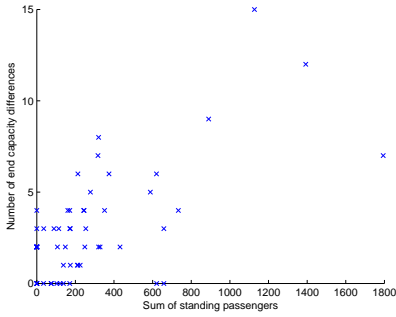
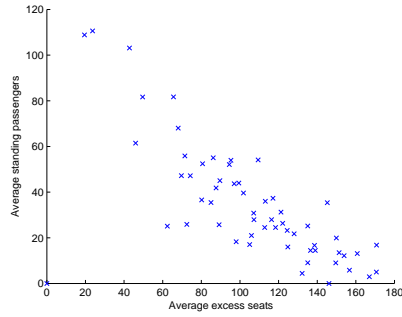
(a) $Weights_1 = (100, 1, 10, 0, 10^4)$ (b) $Weights_2 = (1, 1, 1, 100, 10^5)$

Figure E.10: Sum of standing passengers versus number of end capacity differences.

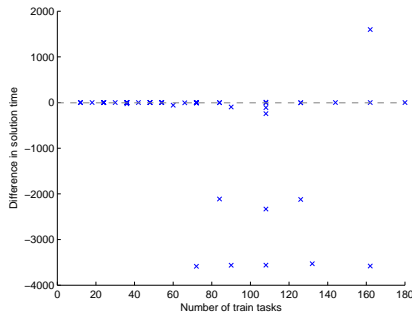


Figure E.11: The difference in computation time ($weights_1 - weights_2$) for each instance versus the number of train tasks.

weight to the number of standing passengers we experience an increase in excess seats.

In practice it is subjective whether emphasis must be on e.g. low number of standing passengers or low end capacity differences on depots. How the weights are set will affect the computation time.

E.9.2 Experimental results for Routing model objective function weights

As for the Position model we have for the Routing model run experimental instances for a set of different weight sets. We have used three setups, see Tab. E.7. The first setup includes the Sequence model solutions in each run of an instance. The second setup discards the Sequence model solution. The third setup varies the use of the Sequence model over including the model, excluding the model or including the model as preferences in the objective function. In the latter case, preferences are generated from the result of the Sequence model. That is, if a train unit has been assigned to a train sequence in the Sequence model, there is in the Routing model a high preference for assigning the same train unit to the train tasks of the train sequence in the Routing model.

For each of the setups presented in Tab. E.7 we have used a factorial design to perform a set of experiments. The factors are the weights in the objective function. The weight of covering a task is named factor A, the weight assigned to sources is named factor B, sinks are named C and the weight of the binary variables telling whether two subsequent train tasks are assigned to the same train unit is named D. In the instances following the third setup the varying use of the Sequence model is included as a factor E. The value levels of each factor used are listed in Tab. E.8.

We run instances based on the *A&A+* and *C* train lines described in Section E.9.1. A full DOE contains $3^5 = 243$ experiments for *Setup*₃ and 81 for *Setup*₁ and *Setup*₂. By using the DOE the number of runs has been reduced for each Setup according to Tab. E.7.

As we are interested in a reasonable solution quality within a short computation time we put an upper limit on the computation time of each run. Prior to each run of the Routing model an execution of the Position model finds the number of train units of each type to assign to each train task. Hereafter, the Sequence model is run. The upper limits on the computation time of the Position model is 600 seconds. The Sequence model is solved at an aggregated level and needs

Number	Setup	Number of runs
<i>Setup₁</i>	Including the Sequence model in each experiment	53
<i>Setup₂</i>	Excluding the Sequence model in each experiment	53
<i>Setup₃</i>	Varying the use of Sequence model as a factor	72

Table E.7: Experimental setups used for the routing parameter choice experiments.

Level	A	B	C	D	E
0	0	0	0	0	Incl. Seq.
1	100	-10	-10	1000	Pref. from Seq.
2	1000	-50	-50	10000	Excl. Seq.

Table E.8: The different levels used for factor experiments of the Routing model.

no upper limit as it always solves to optimality in less than 1 second for the instances chosen in our test setups. Finally, we have set the upper limit on the Routing model computation time to 3600 seconds.

In Tab. E.9 the error contributions in percentage of the average objective are listed for five different measures for each of the experimental setups³. The highest average error contributions are of the tests on *Setup₃* where factor E is included. For *Setup₂* the error contributions are higher for instances on train lines *A&A+* than those on *C*. All average error contributions are high when estimating computation time.

The results suggest that an estimate has a high error if many of the runs in the experiment cannot be solved to optimality. Also, if the use of the sequence model is varied, the error contribution will be high. Even though the average error contributions are low on various objectives of the experimental setup, the average error contribution on the estimate of computation time is high indicating that computation time cannot be predicted with the statistical function. Given these observations we have chosen to use the experiments based on setup 1 from Tab. E.7 to base the choice of weight set used for further experiments.

Given the experiments corresponding to *setup₁* we have filtered the solution data relative to the maximum difference in depot end capacity, the maximum number of standing passengers and the maximum average number of standing passengers. Based on the filtering for train lines *A&A+*, we have chosen the

³For information on factor contributions see Appendix E.C

Objective	<i>Setup₃</i>		<i>Setup₂</i>		<i>Setup₁</i>	
	C	A, A+	C	A, A+	C	A, A+
Standing passengers	45.78	23.85	0.27	5.14	$1.98 \cdot 10^{-13}$	0.93
Excess seats	6.84	4.12	0.38	5.32	$2.49 \cdot 10^{-12}$	0.22
Driven kilometers	7.22	5.00	0.13	3.37	0	0.07
End capacity diff.	4.60	4.85	0.88	4.02	$2.36 \cdot 10^{-12}$	1.59
Computation time	16.38	14.05	16.44	0.01	18.96	15.01

Table E.9: The average error contribution on different objectives given the estimated objective function.

weight set, $Weights_1 = (100, 0, -50, 1000)$, for further experiments. We have chosen not to use the results for line C on $setup_1$ as there is too little difference in the solutions i.e. it is very easy to achieve a good solution. For comparisons we have chosen the weight set, $Weights_2 = (100, -10, -10, 1000)$. We expect that $Weights_2$ will provide the same quality in results as $Weights_1$ as they give similar weights to sinks and sources and the same weights on train tasks and subsequent covers. We want to verify this and to see if there is any difference in computational time.

$Weights_1$ and $Weights_2$ have been used in two separate experiments of 36 runs counting 12 line combinations and three time periods. The line combinations are listed in Tab. E.10. The time periods are all starting at 7 o'clock and are of respectively 1, 2 and 3 hours of duration. In the 36 runs the Sequence model is included in the solution process.

In 5 instances out of the 36 instances a solution for the underlying Position problem could not be found within 600 seconds. We will discard these when evaluating the quality of the Routing model.

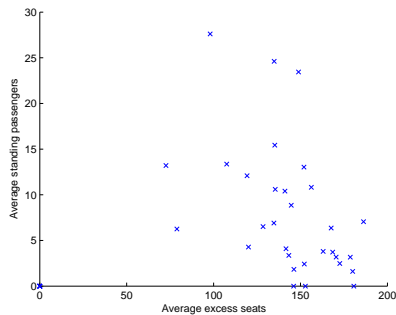
In Fig. E.12 two plots are given of the average excess seats versus the average standing passengers. There is only little difference between the $Weights_1$ solutions in Fig. E.12(a) and the $Weights_2$ solutions in Fig. E.12(b).

Fig. E.13 shows two plots of the sum of standing passengers versus the difference in depot end capacities. As for the plots in Fig. E.12 there is only little difference between the $Weights_1$ solutions in Fig. E.13(a) and the $Weights_2$ solutions in Fig. E.13(b).

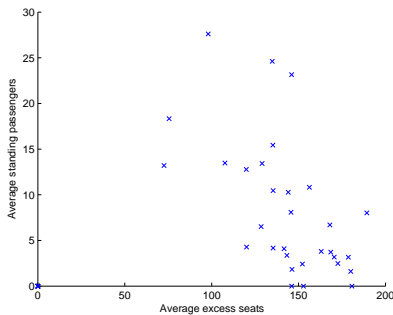
The difference between $Weights_1$ solutions and $Weights_2$ solutions for the Position solution differences is illustrated in Fig. E.14. We see that there is a

Nr.	Lines
1	A
2	C
3	E
4	A, A+
5	C, A+
6	E, A+
7	E, A+, A
8	C, H+, H
9	E, H+, C
10	A+, H, H+, C
11	H+, H, C, A+, A
12	H+, H, C, A+, A, E

Table E.10: Lines included in experiments, see Fig. E.7.



(a) $Weights_1 = (100, 0, -50, 1000)$



(b) $Weights_2 = (100, -10, -10, 1000)$

Figure E.12: Average excess seats versus average standing passengers.

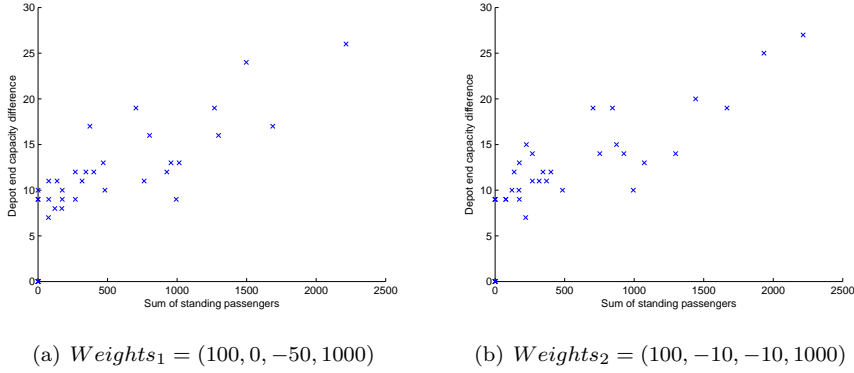


Figure E.13: Sum of standing passengers vs. difference in end capacity.

difference to the Position solution when either the type of train unit assigned in the Routing model does not match the type assigned in the Position model or the number of train units assigned in the Position model does not match the number of train units assigned in the Routing model. We see that there is only little difference in the differences from $Weights_1$ solutions to $Weights_2$ solutions. The average difference over all runs in difference to Position solution is $\Delta Pos_1 = 3.0556$ for $Weights_1$ solutions and $\Delta Pos_2 = 3.0833$ for $Weights_2$ solutions.

Fig. E.15 illustrates the number of train tasks on each run versus the difference in computation time between the $Weights_1$ solutions and the $Weights_2$ solutions. The difference in computation times in the two sets of solutions is with but one exception less than 8 seconds. Considering the computation times of the problem instances only three of these have a computation time higher than 15 seconds. The immediate reason for the deviating results is probably that only few assignments were made in the intermediate step of the Sequence model which of course decreased the number of preassigned variables in the Routing model.

There is a marginal difference in excess seats and standing passengers. In the $Weights_1$ solutions there are slightly fewer standing passengers than the $Weights_2$ solutions. In the $Weights_2$ solutions there are slightly fewer excess seats than the $Weights_1$ solutions. This can all be traced to the difference in deviations from the Position model.

There is only little difference in computation time for the two weight sets. The total average computation time is 215.81 for $Weights_1$ and 214.10 for $Weights_2$.

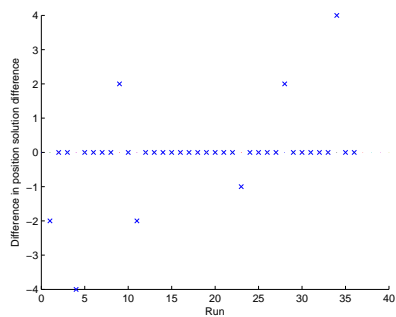


Figure E.14: The difference between $Weights_1$ -solutions and $Weights_2$ -solutions for the Position solution differences.

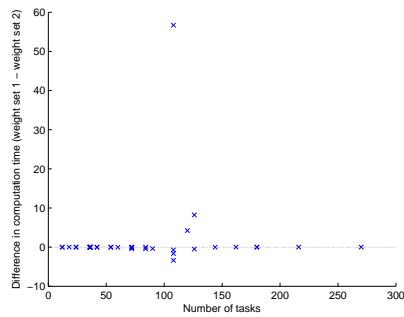


Figure E.15: The number of train tasks versus the difference in solution time for each instance.

Parameter set	Line set	Length of time interval	Opt. gap 60 sec. (%)	Opt. gap 3600 sec. (%)	Opt. gap 28800 sec. (%)
$Weights_1$	9	2	0.88	0.70	0.61
$Weights_1$	10	1	0.82	0.62	0.38
$Weights_2$	9	2	0.90	0.68	0.59
$Weights_2$	10	1	2.01	1.09	0.02

Table E.11: Computational time and optimality gap for 8 hour runs.

Only three of the instances considered have exceedingly high computation times. One of these solve to optimality in 531.25 seconds for $Weights_1$ and 474.56 seconds for $Weights_2$. Discarding these three instances which results in a high computation time the mean computation time is 1.34 seconds for $Weights_1$ and 1.18 seconds for $Weights_2$.

Running the two instances that do not solve to optimality within an hour for a longer period of 8 hours for both $Weights_1$ and $Weights_2$ we get the results in Tab. E.11. We see that increasing the upper time limit on running time does not result in optimal solutions. In fact, the solution quality only improves very little in the 7 hours increased solution time. Hence, a solution close to the optimal solution is obtained within the first 60 seconds for both instances. This indicate that the Routing model even for these instances is practical applicable.

E.9.3 Effect of including the Sequence model

In this section we analyze the 3·36 experiments run for $Weights_1$. Test instances are constructed given the three time windows of 1, 2 and 3 hours starting from 7 o'clock and the line combinations in Tab. E.10. Each of the 36 instances are solved using three different approaches, excluding the Sequence model, including the Sequence model and including the Sequence model as preference in the objective function.

We will in the following refer to the solution approach where the Sequence model solution is included in the Routing solution procedure as $A_{Incl.}$. The solution approach where the Sequence model used as preferences in the Routing solution procedure we refer to as $A_{Pref.}$. Last, the solution approach where we exclude the Sequence model solution we refer to as $A_{Excl.}$.

Fig. E.16 shows respectively the sum of standing passengers for each run for

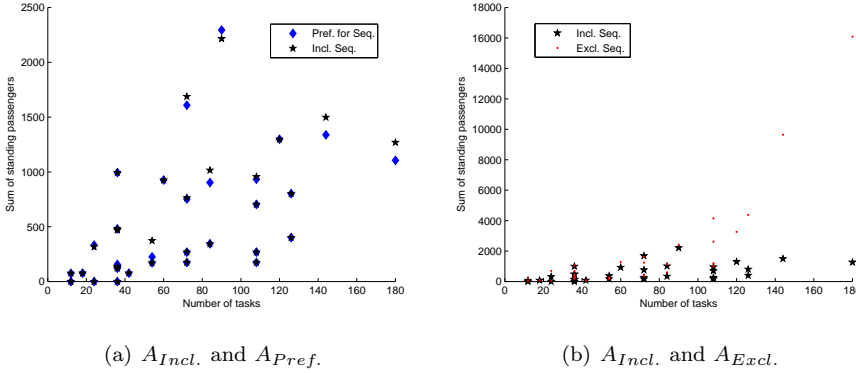


Figure E.16: Sum of standing passengers for each run.

$A_{Incl.}$ & $A_{Pref.}$ and for $A_{Incl.}$ & $A_{Excl.}$. For $A_{Incl.}$ and $A_{Pref.}$, see Fig. E.16(a), the sum of standing passengers are quite close. For $A_{Excl.}$ the sum of standing passengers is in general much higher. Note the difference in scale on the y-axis.

Fig. E.17 illustrates the sum of excess seats for each run for respectively $A_{Incl.}$ & $A_{Pref.}$ and for $A_{Incl.}$ & $A_{Excl.}$. Again the sum of excess seats for $A_{Incl.}$ and $A_{Pref.}$ are close, see Fig. E.17(a). As the sum of excess seats is often a conflicting objective to the sum of standing passengers it is expected that $A_{Excl.}$ has the same or fewer excess seats than $A_{Incl.}$. This is also what we observe in Fig. E.17(b).

The difference in end capacity is illustrated in Fig. E.18. Again, we see that there is a little difference in quality regarding $A_{Incl.}$ and $A_{Pref.}$. When regarding $A_{Excl.}$ the quality decreases.

Fig. E.19 shows the distribution of the results with respect to computation time. $A_{Incl.}$ has the most short running times and only few very high running times. The general mean computation time for $A_{Incl.}$, $A_{Pref.}$ and $A_{Excl.}$ is respectively $\mu_{Incl.} = 250.61$, $\mu_{Pref.} = 510.34$ and $\mu_{Excl.} = 1865.06$.

When for $A_{Incl.}$ disregarding the two runs where an optimal solution can not be found within the 1 hour time limit the mean computation time is $\mu_{Incl.}^{Modified} = 19.62$ seconds. For $A_{Pref.}$ there are three runs where an optimal solution can not be found within the 1 hour time limit. A modified mean time limit is $\mu_{Pref.}^{Modified} = 179.29$.

Summing up the observations we have presented in this section it seems that the

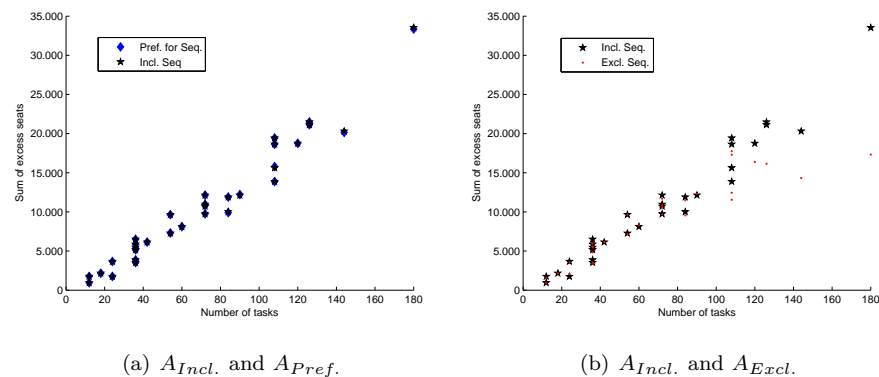


Figure E.17: Sum of excess seats for each run.

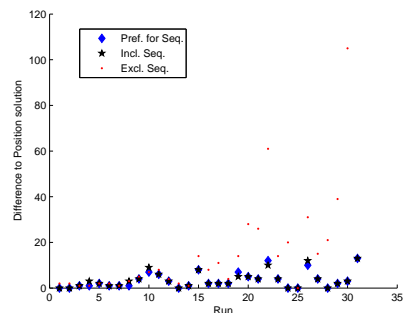


Figure E.18: The difference to the position model solution for each run.

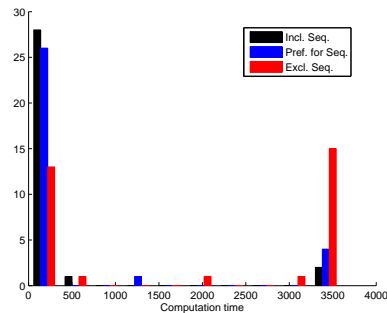


Figure E.19: The distribution of runs with respect to computation time.

solution quality for respectively $A_{Incl.}$ and $A_{Pref.}$ are comparable. The solution quality of $A_{Excl.}$ is lower, most likely because the optimal solution cannot be found within the Routing model running time limit. Even though the $A_{Pref.}$ renders the same solution quality as $A_{Incl.}$ its computation time is on average more than 50% higher. Hence, if we want to obtain an acceptable solution time and quality in real time we must include the Sequence model in the solution process.

In the $A_{Pref.}$ instances there is a higher degree of freedom for assigning values to variables than in the $A_{Incl.}$ instances. However, it is observed that even when $A_{Pref.}$ solves to optimality, the solution quality is at most marginally better on the chosen measures. This indicates that the inclusion of the Sequence model decreases the solution quality only marginally.

E.10 Conclusion

In this paper we have addressed the RSRP. We have formulated a solution approach based on decomposition and consisting of three models to be solved iteratively. The models are implemented with commercial software and initial computational results indicate that the models provide a feasible approach for practical problems up to at least 100 train tasks.

The sequence model is an important step in the solution approach. The average solution time when leaving out the sequence model is 1865.06 seconds. When the Sequence model is included and the variables are locked accordingly the average solution time is 250.61. Furthermore, when the Sequence model is left out the solution quality deteriorates, that is, fewer problem instances solve to optimality within the upper time limit set on computation time.

The quality of solutions when using the Sequence model as preferences compared to locking the variables in the Routing model is the same or only marginally better. We therefore conclude that the Sequence model can be included without deteriorating the solution quality more than marginally. This is desirable as the computation time is decreased 50 % when locking the Routing model variables.

Further research concerns other solution methods for the RSRP. An integrated solution approach may be a heuristic approach solving the Position and Routing problem in one. Also, replacing the Sequence and Routing problems with a column generation approach is interesting.

Appendix E.A Factor contributions, Position model: 6 factors included

In Tab. E.12 the first order factor contributions are listed for the factor experiments on the position model having 6 factors. The factors, A to F, are listed below.

- A** Weight on the term of the sum of excess seats in the Position model objective function.
- B** Weight on the term of the sum train unit kilometers in the Position model objective function.
- C** Weight on the term of the sum of standing passengers in the Position model objective function.
- D** Weight on the term of the sum of couplings in the Position model objective function.
- E** Weight on the term of the sum of depot end capacities in the Position model objective function.
- F** Experimental instance input.

First order contributions are calculated for 6 different measures. For each measure the contribution from a factor is listed for each level higher than 0 e.g. the contribution of A1 to the sum of standing passengers indicates that factor A on the first level higher than zero has a high decreasing effect on the sum of standing passengers, see Tab. E.4.

Standing passengers The sum of standing passengers on all train tasks.

Train unit kilometers The sum of train unit kilometers on all train tasks.

Excess seats The sum of excess seats on all train tasks.

Composition changes The sum of couplings on all train tasks.

End capacity differences The sum of differences to the scheduled end capacity on all depots by the end of recovery.

Computation time

	Standing pass.	Train unit Km	Excess seats	Composition changes	End capacity differences	Comp. time
A1	-1507.3	-199.3	-2981.6	9.1	-0.4	-85.3
A2	-590.4	-292.7	-3638.7	4.3	0.3	76.4
A3	943.3	-340.4	-3621.4	9.9	-2.2	-106.3
B1	3809.3	-795.7	-9364.4	10.4	2.1	41.1
C1	-934.1	19.0	928.7	5.9	5.5	2.8
C2	1304.5	-345.5	-3529.1	-4.4	5.9	168.7
C3	811.4	-374.8	-4224.8	0.5	1.4	116.9
D1	-164.1	-347.6	-2688.1	-1.9	1.4	187.1
D2	919.2	-436.9	-2838.2	-4.8	5.7	178.0
E1	1610.6	-305.7	-1929.3	9.3	-3.4	120.6
E2	2463.5	-520.6	-3896.3	7.6	-1.9	158.5
F1	3042.9	1580.0	5958.7	15.0	6.7	130.2

Table E.12: 1. order factor contributions, 6 factor experiments, Position model

Appendix E.B Factor contributions, Position model:
5 factors included

In Tab. E.13 to E.16 the first order factor contributions are listed for the factor experiments on the position model having 5 factors. The factors, A to E, and the measures are described in Appendix E.10.

	Composition changes	End cap. differences	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	5.7	-1.6	-4500.2	-216.1	33.2	-0.9
A2	4.8	-2.3	-5191.8	-265.1	67.1	-0.8
A3	4.2	-3.3	-4468.2	-234.2	80.1	-0.6
B1	4.6	-0.8	-5068.3	-260.2	61.0	-0.4
C1	1.6	-0.9	-86.7	13.8	-7.5	0.0
C2	2.0	1.5	-802.7	-40.2	-71.9	0.3
C3	4.0	0.8	-1735.3	-83.6	-38.0	-0.3
D1	-1.8	-0.8	-1432.4	-60.6	56.5	0.1
D2	-4.2	1.3	-698.0	-13.5	6.5	0.3
E1	2.2	-2.3	-1504.4	-69.8	14.9	-0.6
E2	1.3	-2.1	-1851.5	-79.1	32.5	-0.7

Table E.13: 1. order factor contributions, 5 factor experiments, Position model, Line C instances

	Composition changes	End cap. differences	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	-2.2	-7.9	1332.6	-62.4	142.6	-71.3
A2	-17.3	-2.5	871.4	44.6	-197.2	-202.0
A3	6.2	-9.7	1171.8	134.0	-807.9	-47.3
B1	2.7	1.2	-242.7	315.0	-6286.1	-102.7
C1	3.3	-8.1	448.2	46.7	-39.2	-204.1
C2	-2.9	-3.7	-1405.9	88.4	-1105.6	-149.4
C3	-9.6	-6.8	2411.2	212.6	-1576.8	-21.5
D1	8.4	0.0	-1798.3	18.0	-1714.9	-158.5
D2	-11.3	4.3	-944.4	17.8	-536.4	-22.6
E1	7.3	-5.8	2109.9	1.9	737.9	52.5
E2	-0.5	-7.4	2239.4	187.2	-1236.2	-177.2

Table E.14: 1. order factor contributions, 5 factor experiments, Position model, line A and A+ instances

	Composition changes	End cap. differences	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	3.7	1.3	-4757.1	-481.4	-75.4	-1.6
A2	5.0	1.8	-6011.4	-662.6	327.2	-2.2
A3	5.5	2.6	-6442.8	-648.4	-27.1	0.5
B1	0.8	0.3	-4159.1	-475.8	254.8	-4.1
C1	0.8	-0.2	-1492.9	-135.2	-378.9	-2.5
C2	-1.5	-0.7	-103.9	36.6	-723.1	-3.6
C3	-1.3	-0.8	870.9	133.0	-823.2	-5.0
D1	1.0	0.3	-1813.3	-232.7	162.9	-3.8
D2	-3.2	-0.2	-1663.3	-214.9	-5.3	-2.2
E1	-3.3	-3.0	-937.4	-87.2	-116.2	5.1
E2	-1.3	-3.3	-1806.6	-234.1	97.8	0.6

Table E.15: 1. order factor contributions, 5 factor experiments, Position model, Line H and H+ instances

	Composition changes	End cap. differences	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	2.7	-0.2	-7436.6	-661.5	498.2	-0.1
A2	1.8	-0.6	-8354.2	-780.0	935.1	-0.1
A3	2.3	-0.4	-7983.8	-734.3	929.9	-0.1
B1	0.5	-0.3	-7642.5	-703.4	852.1	-0.1
C1	1.7	-0.3	-148.5	-6.3	-16.2	0.0
C2	1.5	0.1	1485.6	176.6	-669.8	-0.1
C3	-0.2	-0.6	1638.9	188.9	-510.8	-0.1
D1	-1.9	-0.5	-2598.8	-269.6	574.5	0.0
D2	-2.1	-0.7	-2666.3	-291.4	568.5	0.0
E1	2.4	-2.6	-849.2	-67.5	-36.3	0.0
E2	1.1	-2.4	-1863.5	-162.4	218.9	-0.1

Table E.16: 1. order factor contributions, 5 factor experiments, Position model

Appendix E.C Factor contributions, Routing model

In Tab. E.17 to E.22 the first order factor contributions are listed for the factor experiments on the Routing model. The factor experiments relative to Tab. E.17 to E.18 have 5 factors A to E. The factor experiments relative to Tab. E.19 to E.22 have 4 factors A to D. The factors are described below. Levels of the factors are described in Tab. E.8.

- A** Weight on the term of the tasks in the Routing model objective function.
- B** Weight on the term of the sources in the Routing model objective function.
- C** Weight on the term of the sinks in the Routing model objective function.
- D** Weight on the term of the consecutive covered tasks in the Routing model objective function.
- E** Use of the train Sequence model.

The first order contributions are calculated for 6 different measures.

Difference to Position solution The sum of assignments made in the Routing solution which differs from the assignments in the Position solution.

Difference to end capacity The sum of differences to the scheduled end capacity on all depots by the end of recovery.

Excess seats The sum of excess seats on all train tasks.

Train unit kilometers The sum of train unit kilometers on all train tasks.

Standing passengers The sum of standing passengers on all train tasks.

Computation time

	Difference to Pos. solution	Diff. to end cap.	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	1,4	0,4	-222,9	-27,4	293,2	0,0
A2	11,6	3,7	-1735,6	-205,1	2175,9	0,1
B1	-9,3	-2,5	1384,4	153,6	-1306,8	0,1
B2	-8,9	-2,4	1315,1	145,2	-1217,2	0,0
C1	-3,0	-0,8	520,6	66,6	-716,3	0,0
C2	-1,8	-0,4	327,6	43,0	-464,2	0,0
D1	-3,0	-0,8	499,1	63,1	-677,4	0,0
D2	-2,4	-0,6	415,5	53,7	-578,5	0,0
E1	-8,1	-2,1	1188,5	129,6	-1051,1	0,0
E2	-7,9	-2,1	1148,9	124,7	-998,2	0,0

Table E.17: 1. order factor contributions, factor experiments, Routing model, Line C instances, Choice of Sequence model

	Difference to Pos. solution	Diff. to end cap.	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	3,3	0,5	-308,9	-89,9	1021,7	-22,3
A2	41,1	12,4	-4370,6	-822,7	10348,4	144,1
B1	-46,6	-6,5	4835,9	950,1	-8807,7	89,9
B2	-44,7	-5,1	4810,2	919,8	-8575,5	81,8
C1	-4,1	-1,3	364,3	154,6	-1822,4	51,0
C2	-2,3	-0,6	108,8	83,6	-1263,3	33,9
D1	-7,9	-0,3	738,7	173,4	-2553,3	50,4
D2	-1,5	-1,0	8,1	76,4	-1009,2	42,9
E1	-44,0	-3,8	4459,4	893,2	-8058,4	70,0
E2	-43,7	-5,2	4533,3	873,6	-8062,8	65,8

Table E.18: 1. order factor contributions, factor experiments, Routing model, Line A and A+ instances, Choice of Sequence model

	Difference to Pos. solution	Diff. to end cap.	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	0,2	-6,4	3579,0	426,4	-4227,3	0,2
A2	0,2	-6,4	3587,0	426,8	-4229,7	0,2
B1	0,2	-0,4	64,0	2,6	-18,7	0,2
B2	0,1	-0,4	56,0	2,3	-16,3	0,1
C1	0,0	0,0	0,0	0,0	0,0	0,0
C2	-0,1	0,2	-32,0	-1,3	9,3	-0,1
D1	0,2	-6,4	3579,0	426,4	-4227,3	0,2
D2	0,2	-6,4	3587,0	426,8	-4229,7	0,2

Table E.19: 1. order factor contributions, factor experiments, Routing model, Line C instances, No use of Sequence model

	Difference to Pos. solution	Diff. to end cap.	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	-68,3	-6,3	6909,6	1616,3	-18028,4	299,6
A2	-66,3	-8,4	7119,2	1691,5	-18514,8	299,5
B1	-2,3	-2,4	76,1	21,1	113,1	0,0
B2	1,0	-2,3	-30,2	16,2	0,8	0,0
C1	3,4	-0,7	-271,1	-59,9	459,3	0,0
C2	2,3	0,6	-189,9	-46,7	277,1	0,1
D1	-71,8	-9,4	7387,9	1718,0	-18662,1	299,5
D2	-72,0	-10,4	7465,6	1665,5	-18739,4	299,5

Table E.20: 1. order factor contributions, factor experiments, Routing model, Line A and A+ instances, No use of Sequence model

	Difference to Pos. solution	Diff. to end cap.	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	-4,0	-1,0	494,0	41,6	-106,0	0,0
A2	-4,0	-1,0	494,0	41,6	-106,0	0,0
B1	0,0	0,0	0,0	0,0	0,0	0,0
B2	0,0	0,0	0,0	0,0	0,0	0,0
C1	0,0	0,0	0,0	0,0	0,0	0,0
C2	0,0	0,0	0,0	0,0	0,0	0,0
D1	-4,0	-1,0	494,0	41,6	-106,0	0,0
D2	-4,0	-1,0	494,0	41,6	-106,0	0,0

Table E.21: 1. order factor contributions, factor experiments, Routing model, Line C instances, Use of Sequence model

	Difference to Pos. solution	Diff. to end cap.	Excess seats	Train unit Km	Standing pass.	Comp. time
A1	-38,0	-5,1	4079,4	681,9	-5340,6	0,6
A2	-38,0	-3,4	4062,2	674,0	-5357,8	0,6
B1	0,0	0,1	53,4	-4,0	53,4	0,4
B2	0,0	0,1	20,1	-3,2	20,1	0,3
C1	0,0	-0,2	-32,7	-1,6	-32,7	0,1
C2	0,0	-0,2	-40,8	0,8	-40,8	-0,1
D1	-38,0	-2,8	3972,8	670,0	-5447,2	0,7
D2	-38,0	-3,7	3973,6	669,2	-5446,4	0,5

Table E.22: 1. order factor contributions, factor experiments, Routing model, Line A and A+ instances, Use of Sequence model

Bibliography

- A. Alfieri, R. Groot, L. Kroon, and A. Schrijver. Efficient circulation of railway rolling stock. *Transportation Science*, 40(3):378–391, Aug. 2006.
- J. Kintanar C. Queille W. Stripling Ben-Khedher, N. Schedule optimization at snf: From conception to day of departure. *Interfaces*, 28:6–23, 1998.
- P. Brucker, J. Hurink, and T. Rolfes. Routing of railway carriages. *Journal of Global Optimization*, 27(2-3):313–332, Nov. 2004.
- J.-F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32:380–404, 1998.
- J.-F. Cordeau, G. Desaulniers, N. Lingaya, F. Soumis, and J. Desrosiers. Simultaneously locomotive and car assignment at via rail canada. *Transportation Science*, 35(8):767–787, Sept. 2001.
- P.-J. Fioole, L. G. Kroon, G. Maróti, and A. Schrijver. A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research*, 174:1281–1297, 2006.
- M. Hofman, L. F. Madsen, J. J. Groth, J. Clausen, and J. Larsen. Robustness and recovery in train scheduling - a simulation study from dsb s-tog a/s. In *ATMOS 2006. Informatics and Mathematical Modelling*, Technical University of Denmark, DTU, 2005. IMM-Technical Report-2006-12.
- D. Huisman, L. G. Kroon, R. M. Lentink, and M. J. C. M. Vromans. Operations Research in passenger railway transportation. *Statistica Neerlandica*, 59(4): 467–497, 2005.

- N. Lingaya, J.-F. Cordeau, G. Desaulniers, J. Desrosiers, and F. Soumis. Operational car assignment at via rail canada. *Transportation Science*, 36(9): 755–778, Nov. 2002.
- L. K. Nielsen. A decision support framework for rolling stock rescheduling. Technical report, ARRIVAL Project, August 2008.
- M. Peeters and L. Kroon. Circulation of railway rolling stock: A branch-and-price approach. Technical report.
- N. J. Rezanova and D. Ryan. The train driver recovery problem a set partitioning based model and solution method. *Accepted for publication in Computers and OR*, 2006.
- A. Schrijver. Minimum circulation of railway stock. *CWI Quart*, 6:205–217, 1993.
- J. Törnquist. Computer-based decision support for railway traffic scheduling and dispatching: a review of models and algorithms. In *Presented at ATMOS2005, Palma de Mallorca, Spain, October 2005 (proceedings to be published on the Dagstuhl Research Online Publication Server (DROPS) during 2006*, 2006.
- C. G. Walker, J. N. Snowdon, and D. M. Ryan. Simultaneous disruption recovery of a train timetable and crew roster in real time. *Computers & Operations Research*, 32(8):2077–2094, Aug. 2004.

Bibliography

- E. Abbink, M. Fischetti, L. G. Kroon, G. Timmer, and M. J. C. M. Vromans. Reinventing crew scheduling at netherlands railways. *Interfaces*, 35(5):393–401, 2005.
- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows – Theory, Algorithms and Applications*. Prentice Hall, 1993.
- R. K. Ahuja, Ö. Ergun, J. B. Orlin, and A. P. Punnen. A survey of very large-scale neighborhood search techniques. *Discrete Applied Mathematics*, 123:75–102, 2002.
- R. K. Ahuja, C. B. Cunha, and G. Sahin. *Tutorials in Operations Research*, chapter Network Models in Railroad Planning and Scheduling, pages 54–101. 2005a.
- R. K. Ahuja, J. Liu, J. B. Orlin, D. Sharma, and L. A. Shughart. Solving real-life locomotive-scheduling problems. *Transportation Science*, 39(4):503–517, 2005b.
- A. Alfieri, R. Groot, L. Kroon, and A. Schrijver. Efficient circulation of railway rolling stock. *Transportation Science*, 40(3):378–391, Aug. 2006.
- A. C. Andersen, C. G. Christensen, I. L. T. Heilmann, and C. T. Holst. Planlægning af køreplaner. Bachelor project, The Technical University of Denmark, Department of Informatics and Mathematical Modeling, 2006.
- Y. P. Aneja, V. Aggarwal, and K. P. K. Nair. Shortest chain subject to side constraints. *Networks*, 13:295–302, 1983.

- N. Ben-Khedher, J. Kintanar, C. Queille, and W. Stripling. Schedule optimization at sncf: From conception to day of departure. *Interfaces*, 28:6–23, 1998.
- L. Bengtsson, R. Galia, T. Gustafsson, C. Hjorring, and N. Kohl. *Railway Optimization 2004*, volume 4359 of *Lecture Notes in Computer Science*, chapter Railway Crew Pairing Optimization, pages 126–144. Springer–Verlag Berlin Heidelberg, 2007.
- N. Boland, J. Dethridgea, and I. Dumitrescu. Accelerated label setting algorithms for the elementary resource constrained shortest path problem. *Operations Research Letters*, 32(8):58–68, Jan. 2006.
- R. Borndörfer, M. Grötschel, and M. E. Pfetsch. *Computer-aided Systems in Public Transport*, chapter Models for Line Planning in Public Transport, pages 363–378. Springer Berlin Heidelberg, 2008.
- R. Borndörfer, M. Grötschel, and M. E. Pfetsch. A column-generation approach to line planning in public transport. *Transportation Science*, 41(1):123–132, 2007.
- P. Brucker, J. Hurink, and T. Rolfes. Routing of railway carriages. *Journal of Global Optimization*, 27(2–3):313–332, Nov. 2004.
- M. R. Bussieck. *Optimal Lines in Public Rail Transport*. PhD thesis, Technical University Braunschweig, Braunschweig, 1998.
- M. R. Bussieck, T. Lindner, and M. E. Lübbecke. A fast algorithm for near optimal line plans. *Mathematical Methods of Operations Research*, 59(2):205–220, 2004.
- A. Caprara, L. Kroon, M. Monaci, M. Peeters, and P. Toth. *Handbook in OR & MS*, volume 14, chapter Passenger Railway Optimization, pages 129–187. Elsevier, 2007.
- J. Clausen, A. Larsen, and J. Larsen. Disruption management in the airline industry – concepts, models and methods. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2005.
- J.-F. Cordeau, G. Desaulniers, N. Lingaya, F. Soumis, and J. Desrosiers. Simultaneously locomotive and car assignment at via rail canada. *Transportation Science*, 35(8):767–787, Sept. 2001.
- A. D’Ariano. *Improving Real-Time Train Dispatching: Models, Algorithms and Applications*. PhD thesis, The Technical University of Delft, 2008.
- J. Desrosiers and M. E. Lübbecke. Selected topics in column generation. *Operations Research*, 53(6):1007–1023, 2005.

- A. P. Goldie S. Miller J. A. Meyer D. M. Ryan A. C. Scott E. R. Butchers, P. R. Day and Ch. A. Wallace. Optimized crew scheduling at air new zealand. *Interfaces*, 31(1):30–56, 2001.
- P.-J. Fioole, L. G. Kroon, G. Maróti, and A. Schrijver. A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research*, 174:1281–1297, 2006.
- M. Folkmann, J. Jespersen, and M. N. Nielsen. *Algorithmic Methods for Railway Optimization*, volume 4359, chapter Estimates on rolling stock and crew in DSB S-tog based on timetables, pages 91–107. Springer, 2007.
- P. Føns. Decision support for depot planning in the railway industry. Master’s thesis, Departments of Informatics and Mathematical Modeling, The Technical University of Denmark, Richard Petersens plads, DK-2800 Lyngby, 2006. Supervised by Jesper Larsen.
- R. Freling, R. M. Lentink, L. G. Kroon, and D. Huisman. Shunting of passenger train units in a railway station. *Transportation Science*, 39:261–272, 2005.
- J. H. M. Goosens. *Models and algorithms for railway line planning problems*. PhD thesis, University of Maastricht, Maastricht, 2004.
- J. H. M. Goosens, C. P. M. Van Hoesel, and L. G. Kroon. A branch-and-cut approach for solving railway line planning problems. *Transportation Science*, 38:379–393, 2004.
- R.M.P. Goverde and M.A. Odijk. *Computers in Railways*, volume VIII, chapter Performance Evaluation of Network Timetables Using PETER, pages 731–740. Southampton, UK: WIT Press, 2002.
- J. J. Groth, J. Clausen, and J. Larsen. Optimal reinsertion of cancelled train line. Technical report, The Technical University of Denmark, Department of Management Engineering, 2006.
- J. J. Groth, D. Potthoff, J. Clausen, D. Huisman, L. Kroon, G. Maróti, and M. N. Nielsen. Disruption management in passenger railway transportation. Technical report, The Technical University of Denmark, Department of Management Engineering, 2007.
- J. J. Groth, J. Larsen, and J. Clausen. The rolling stock recovery problem. Technical report, The Technical University of Denmark, Department of Management Engineering, 2008.
- M. Hofman, L. F. Madsen, J. J. Groth, J. Clausen, and J. Larsen. Robustness and recovery in train scheduling – a simulation study from dsb s-tog a/s. In *ATMOS 2006*. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2006. IMM-Technical Report-2006-12.

- D. Huisman, L. G. Kroon, R. M. Lentink, and M. J. C. M. Vromans. Operations Research in passenger railway transportation. *Statistica Neerlandica*, 59(4): 467–497, 2005.
- P. M. Jacobsen. Rangerplanlægning. Master’s thesis, Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen Ø, 2008. Supervised by David Pissinger, DIKU.
- A.S. Jain and S. Meeran. A state-of-the-art review of job-shop scheduling techniques. *European Journal of Operational Research*, 113:390–434, 1999.
- T. S. Jensen. Planlægning af reparationer for dsb s-tog. Master’s thesis, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, 2008. Supervised by Jørgen Bang-Jensen, IMADA.
- N. Kohl and S. E. Karisch. Airline crew rostering: Problem types, modeling, and optimization. *Annals of Operations Research : Paperbound Edition*, 127: 223–259, 2004.
- N. Kohl, A. Larsen, J. Larsen, A. Ross, and S. Tiourine. Airline disruption management—perspectives, experiences and outlook. *Journal of Air Transport Management*, 13:149–162, 2007.
- L. Kroon, G. Maróti, M. R. Helmrich, M. Vromans, and R. Dekker. Stochastic improvement of cyclic railway timetables. *Transportation Research Part B*, 42:553–570, 2008.
- L. G. Kroon and M. Fischetti. *Computer-Aided Scheduling of Public Transport*, chapter Crew scheduling for Netherlands Railways Destination Customer, pages 181–201. Springer, 2001.
- C. Liebchen. *Operations Research Proceedings 2006*, volume 2006, chapter Periodic Timetable Optimization in Public Transport, pages 29–36. Springer Berlin Heidelberg, 2007.
- C. Liebchen and R. H. Möhring. *Algorithmic Methods for Railway Optimization*, chapter The Modeling Power of the Periodic Event Scheduling Problem: Railway Timetables – and Beyond. Springer Berlin Heidelberg, 2007.
- N. Lingaya, J.-F. Cordeau, G. Desaulniers, J. Desrosiers, and F. Soumis. Operational car assignment at via rail canada. *Transportation Science*, 36(9): 755–778, Nov. 2002.
- G. Marquis M. Gamache, F. Soumis and J. Desrosiers. A column generation approach for large-scale aircrew rostering problems. *Operations Research Management Science*, 47(2):247–263, 1999.

- G. Maróti and L. G. Kroon. Maintenance routing for train units: The transition model. *Transportation Science*, 39(4):518–525, 2005.
- G. Maróti and L. G. Kroon. Maintenance routing for train units: The interchange model. *Computers & Operations Research*, 34:1121–1140, 2007.
- D. Middelkoop and M. Bouwman. Simone: Large scale train network simulations. In *Proceeding of the 2001 Winter Simulation Conference*, volume 2, pages 1042–1047, 2001.
- D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, Inc., 5th edition, 1997.
- L. K. Nielsen. Planlægning af arbejdsplaner for togrevisorer i s-toge. Master's thesis, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, 2006. Supervised by Jørgen Bang-Jensen, IMADA.
- L. K. Nielsen. A decision support framework for rolling stock rescheduling. Technical report, ARRIVAL Project, August 2008.
- M. N. Nielsen and T. Christensen. Duty planning and design of experiments. In *ATMOS 2006*, 2006.
- M. N. Nielsen, J. Clausen, and B. Hove. Constructing periodic timetables using mip – a case study from dsb s-train. *International Journal of Operational Research*, 1(3):213–227, 2006.
- The Danish Ministry of Transport. Kontrakt mellem trafikministeriet og dsb stog a/s om s-togs trafikken udført som offentlig service i perioden 2005–2014. Available online, 2004.
- M. Peeters and L. Kroon. Circulation of railway rolling stock: A branch-and-price approach. Technical report, ERIM, Erasmus University Rotterdam, 2003.
- L. D. Pietrantonio and J. Pelkmans. The economics of eu railway reform. College of Europe, Bruges European Economic Policy Briefings, Sep. 2004. <http://www.coleurop.be/content/studyprogrammes/eco/publications/...BEEPs/BEEP8.pdf>.
- N. J. Rezanova and D. Ryan. The train driver recovery problem – a set partitioning based model and solution method. *Accepted for publication in Computers and OR*, 2006.
- D. M. Ryan and J. C. Falkner. On the integer properties of scheduling set partitioning models. *European Journal of Operational Research*, 35(3):442–456, 1988.

- A. Schrijver. Minimum circulation of railway stock. *CWI Quart*, 6:205–217, 1993.
- P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581, 1989.
- J. C. Villumsen. Construction of timetables based on periodic event scheduling. Master’s thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK–2800 Kgs. Lyngby, 2006. Supervised by Jens Clausen, IMM.
- C. G. Walker, J. N. Snowdon, and D. M. Ryan. Simultaneous disruption recovery of a train timetable and crew roster in real time. *Computers & Operations Research*, 34(1):2077–2094, Aug. 2004.
- White-paper. A strategy for revitalising the community’s railways. EU Commission White Paper, Jul. 1996.